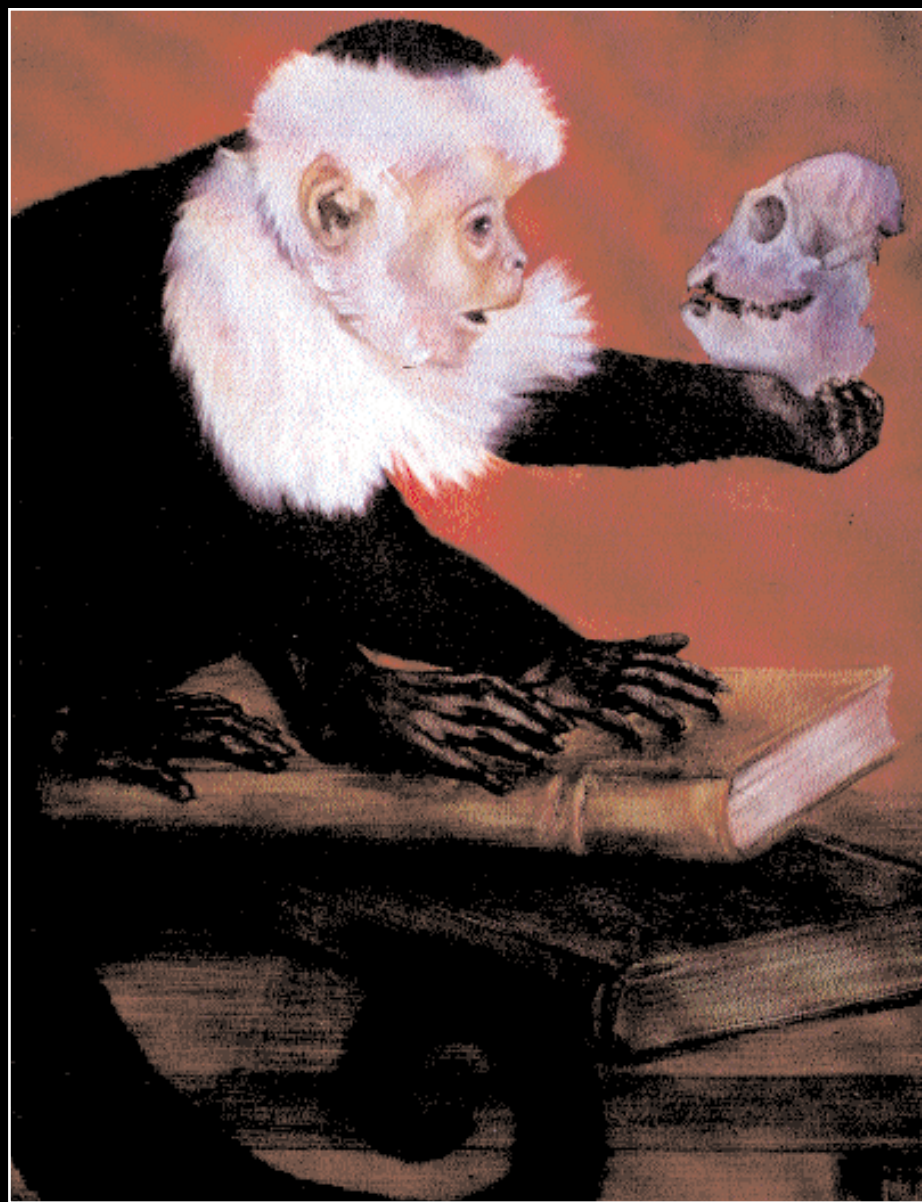


# The Nature of Truth

Classic and Contemporary Perspectives



edited by Michael P. Lynch

---

# The Nature of Truth



---

# The Nature of Truth

Classic and Contemporary Perspectives

edited by

Michael P. Lynch

A Bradford Book

The MIT Press

Cambridge, Massachusetts

London, England

© 2001 by Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, and information storage and retrieval) without permission in writing from the publisher.

Chapter 24, "A Defense of Minimalism," © 2001 by Paul Horwich.

This book was set in Sabon on '3B2' by Asco Typesetters, Hong Kong, and was printed and bound in the United States of America.

First printing, 2001

Library of Congress Cataloging-in-Publication Data

The nature of truth : classic and contemporary perspectives / edited by  
Michael P. Lynch.

p. cm.

"A Bradford Book."

Includes bibliographical references and index.

ISBN 0-262-12239-1 (alk. paper) — ISBN 0-262-62145-2 (pbk. : alk. paper)

1. Truth. I. Lynch, Michael P. (Michael Patrick), 1966—

BD171 .N35 2001

121—dc21

00-048035

For my parents



---

# Contents

Preface xi

Acknowledgments xiii

**Introduction: The Mystery of Truth 1**

**I Realism and the Correspondence Theory 7**

**1 Truth and Falsehood 17**

Bertrand Russell

**2 Truth 25**

J. L. Austin

**3 A Realist Conception of Truth 41**

William P. Alston

**4 Contextual Semantics and Metaphysical Realism: Truth as Indirect Correspondence 67**

Terence Horgan

**II Coherence Theories 97**

**5 Coherence as the Nature of Truth 103**

Brand Blanshard

**6 The Coherence Theory 123**

Ralph C. S. Walker

**7 The Case for Coherence 159**

Linda Martín Alcoff

**III Pragmatism and Verificationism 183**

**8 How to Make Our Ideas Clear 193**

Charles Sanders Peirce



9	Pragmatism's Conception of Truth	211
	William James	
10	Truth	229
	Michael Dummett	
11	Two Philosophical Perspectives	251
	Hilary Putnam	
12	Is Truth a Goal of Inquiry? Donald Davidson versus Crispin Wright	259
	Richard Rorty	
IV	Phenomenological and Postmodernist Conceptions	287
13	On the Essence of Truth	295
	Martin Heidegger	
14	Truth and Power	317
	Michel Foucault	
V	Tarski's Theory and Its Importance	321
15	The Semantic Conception of Truth and the Foundations of Semantics	331
	Alfred Tarski	
16	Tarski's Theory of Truth	365
	Hartry Field	
17	What Is a Theory of Truth?	397
	Scott Soames	
VI	Deflationary Views and Their Critics	419
18	The Nature of Truth	433
	Frank Plumpton Ramsey	
19	Truth	447
	P. F. Strawson	
20	Truth	473
	W. V. O. Quine	
21	Correspondence Truth, Disquotational Truth, and Deflationism	483
	Hartry Field	

22	The Prosentential Theory: Further Reflections on Locating Our Interest in Truth	505
	Dorothy Grover	
23	A Critique of Deflationism	527
	Anil Gupta	
24	A Defense of Minimalism	559
	Paul Horwich	
25	The Metaphysics of Truth	579
	Michael Devitt	
VII	Primitivism, Identity Theory, and Alethic Pluralism	613
26	The Folly of Trying to Define Truth	623
	Donald Davidson	
27	Epistemology and Primitive Truth	641
	Ernest Sosa	
28	Truth: The Identity Theory	663
	Jennifer Hornsby	
29	Truth as Identity and Truth as Correspondence	683
	Marian David	
30	The Face of Cognition	705
	Hilary Putnam	
31	A Functionalist Theory of Truth	723
	Michael P. Lynch	
32	Minimalism, Deflationism, Pragmatism, Pluralism	751
	Crispin Wright	
	Contributors	789
	Index	793



---

## Preface

The deepest philosophical questions are not isolated; they sit at the center of our broader cultural concerns. This is certainly the case with the problem of truth. An increasing philosophical preoccupation with truth over the last hundred years is deeply intertwined with two larger issues. The first issue is pluralism. The sheer number and variety of viewpoints we encounter on any question is forcing us, on both the political and philosophical fronts, to think about how objectivity is possible. The second issue is our increasing technological sophistication at both pursuing and distorting the truth. There is little prospect that we will slow down in either respect at the beginning of this century. In a world where things move so fast that the real can be difficult to tell from the virtual, understanding truth seems more relevant than ever.

This volume is a comprehensive survey of the various attempts to solve this problem. Roughly speaking, the essays center around two questions: Does truth have an underlying nature? And if so, what sort of nature does it have? The book is therefore concerned with the question of truth itself, as opposed to the relation of truth to other issues of philosophical interest, such as knowledge, meaning, and logic. This is the first of the ways in which I've attempted to make the territory more manageable for a single volume. The second is by limiting the essays included to those written during the twentieth century.

The problem of truth is complex, and my hope is that this book will act as a map not only for undergraduate and graduate students of philosophy but also for anyone who finds himself lost in the thickets of the contemporary debate. To this end, the introductions to each part are intended to help the reader locate the most important concepts and issues

discussed by the essays of that part. Of course, like any map, this one is limited by its size and scale. There are a number of deserving essays that could not be included because of the limitations of space. I have tried to address this issue by including suggestions for further reading at the end of every introduction. The reader is strongly encouraged to consult these for a more in-depth look at the various theories.

A word about how the essays are related to each other. I have organized them by theory, but since many of the essays could fit under more than one category, there are other paths to follow through the territory than what I have laid out. I have tried to take account of this fact in the introduction to each part by discussing not only the chapters within the part but also those that appear in other parts of the book yet deal with the issues being discussed in the part in question.

Many people have helped me with this book. Thanks first go to my editorial assistant Sam Hughes, whose sharp thinking and capacity for organization were of invaluable assistance, particularly during the crucial first stages of the book's development. Several people commented on the various introductions: Bob Barnard, Paul Bloomfield, Charles Fletcher, Rex Gilliland, and Thomas Nenon in particular; while conversations with Andrew Cortens, Marion David, Eric Olson, Bill Alston, Terry Horgan, Mark Lance, and others helped me to decide what to include. My colleague William Lawhead provided helpful advice (and a sympathetic ear) and allowed me to borrow a continuous stream of books. The Masters of St. Edmund's College, Cambridge, were gracious enough to provide me with the position of Visiting Scholar in the summer of 1999 and thereby allow me access to one of the world's great libraries. My students at the University of Mississippi over the last few years have tolerated my continual obsession with the nature of truth, and interactions with them have taught me much about it; I am indebted to them one and all. Thanks to Alan Thwaites and the editorial staff at the MIT Press for their expert assistance. Most important, I thank Terry, best friend, best critic, loving partner.

---

# Acknowledgments

## Realism and the Correspondence Theory

Bertrand Russell. “Truth and Falsehood.” Chapter 12 of *The Problems of Philosophy*. Oxford: Oxford University Press, 1912. Reprinted with permission of the publisher.

J. L. Austin. “Truth.” *Proceedings of the Aristotelian Society*, suppl. vol. 24. Reprinted by courtesy of the editor of the Aristotelian Society. Copyright 1950.

## Coherence Theories

Brand Blanshard. “Coherence as the Nature of Truth.” From *The Nature of Thought*, vol. 2, pp. 260–279. 1939. Reprinted with the permission of HarperCollins Publishers.

Ralph C. S. Walker. “The Coherence Theory.” From *The Coherence Theory of Truth*, pp. 1–6 and 25–40. Copyright 1989 by Routledge Press. Reprinted with the permission of Taylor and Francis Books. Section III contains new material added by the author.

## Pragmatism and Verificationism

Charles Sanders Peirce. “How to Make Our Ideas Clear.” *Popular Science Monthly* 12 (1878): 286–302.

William James. “Pragmatism’s Conception of Truth.” From *Pragmatism: A New Name for Some Old Ways of Thinking*, pp. 197–236. New York: Longmans, 1907.

Michael Dummett. "Truth." *Proceedings of the Aristotelian Society* 59. Copyright 1958–1959. Reprinted by courtesy of the editor of the Aristotelian Society.

Hilary Putnam. "Two Philosophical Perspectives." From *Reason, Truth, and History*, pp. 49–56. Copyright 1981 by Cambridge University Press. Reprinted with the permission of Cambridge University Press.

Richard Rorty. "Is Truth a Goal of Inquiry? Donald Davidson versus Crispin Wright." *Philosophical Quarterly* 45 (1995): 281–300. Copyright 1995 by the editors of *The Philosophy Quarterly*. Reprinted with permission.

### Phenomenological and Postmodernist Conceptions

Martin Heidegger. "On the Essence of Truth." From *Basic Writings by Martin Heidegger*. English translation copyright 1977 by Harper and Row Publishers. General Introduction and the introductions to each part copyright by D. F. Krell. Reprinted by permission of HarperCollins Publishers.

Michel Foucault. "Truth and Power." From *Power/Knowledge*, pp. 131–133. Copyright 1972, 1975, 1976, 1977 by Michel Foucault. *Power/Knowledge* copyright 1980 by the Harvester Press. Reprinted by permission of Pantheon Books, a division of Random House.

### Tarski's Theory and Its Importance

Alfred Tarski. "The Semantic Conception of Truth and the Foundations of Semantics." *Philosophy and Phenomenological Research* 4 (1944). Copyright 1992 by the Estate of Alfred Tarski. Reprinted by permission of Jan Tarski.

Hartry Field. "Tarski's Theory of Truth." *Journal of Philosophy* 69, no. 13: 347–375. Reprinted with the permission of the author and *The Journal of Philosophy*.

Scott Soames. "What Is a Theory of Truth?" *Journal of Philosophy* 81, no. 8: 411–429. Reprinted with the permission of the author and *The Journal of Philosophy*.

## Deflationary Views and Their Critics

Frank Plumpton Ramsey. "The Nature of Truth." *Episteme* 16: 6–16. Copyright 1990. Reprinted by permission of Kluwer Academic Publishers.

P. F. Strawson. "Truth." *Proceedings of the Aristotelian Society*, suppl. vol. 24. Copyright 1950. Reprinted by courtesy of the editor of the Aristotelian Society.

W. V. O. Quine. "Truth." From *The Pursuit of Truth*, pp. 77–88. Cambridge: Harvard University Press. Copyright 1990, 1992, by the President and Fellows of Harvard College. Reprinted by permission of the author and the publisher.

Hartry Field. "Correspondence Truth, Disquotational Truth, and Deflationism." Excerpted from "The Deflationary Conception of Truth," in *Fact, Science, and Morality: Essays on A. J. Ayer's Language, Truth, and Logic*, edited by G. Macdonald and C. Wright. Oxford: Blackwell, 1986. Reprinted, with changes by the author, with permission of the author and the publisher.

Anil Gupta. "A Critique of Deflationism." *Philosophical Topics* 21 (1993), no. 2: 57–81. Reprinted by permission of the author and the editor of *Philosophical Topics*.

Michael Devitt. "The Metaphysics of Truth." This essay is also appearing in *What Is Truth?* edited by Richard Schantz. Berlin and New York: Walter de Gruyter, 2000. It is printed here with the permission of Richard Schantz and the author.

## Primitivism, Identity Theory, and Alethic Pluralism

Donald Davidson. "The Folly of Trying to Define Truth." *Journal of Philosophy* 93, no. 6: 263–279. Copyright 1999 by Donald Davidson. Reprinted with permission of the author.

Jennifer Hornsby. "Truth: The Identity Theory." Reprinted with omissions from *Proceedings of the Aristotelian Society* 97. Copyright 1997. Reprinted by courtesy of the author and the editor of the Aristotelian Society.



Hilary Putnam. “The Face of Cognition.” Presented as part of the Dewey Lectures of 1994, “Sense, Nonsense, and the Senses: An Inquiry into the Powers of the Human Mind,” lecture 3: “The Face of Cognition.” Published in *The Journal of Philosophy* 91, no. 9: 488–504 and 510–517. Reprinted with permission of the author and *The Journal of Philosophy*.

Note: In preparing previously published essays for this volume, I have engaged in minor editing.

---

# The Nature of Truth



---

# Introduction: The Mystery of Truth

What is truth?

Pontius Pilate

Humanly speaking, let us define truth, while waiting for a better definition, as—a statement of the facts as they are.

Voltaire

In court, witnesses swear to tell the truth, the whole truth, and nothing but the truth. One is expected to know what this means, and in some sense, it is clear that we do. Yet at the same time, truth seems so stubbornly abstract that, like Pontius Pilate, we treat questions about its nature as rhetorical. We cowardly avoid it, courageously pursue it, and lament its distortion, but when pressed to say what truth is, we find ourselves tongue-tied and frustrated. The nature of truth seems a mystery.

There are some obvious and not so obvious reasons for this fact. The most obvious is the ambiguity of the word. Even if we restrict ourselves to the adjective “true,” one can speak of “true friends,” “true north,” “aiming true,” and so on. The sense of the word that concerns philosophers, however, is the sense being assumed in the very first sentence of this introduction. In the courtroom, we want the witness to speak the truth, to report what she believes to be true, i.e., true propositions. This is the sense of the word that matters most in our everyday lives.

Limiting the scope of the question in this way helps somewhat, but not much. As Voltaire’s droll remark illustrates, it may seem as if one can define truth only by platitudes, by saying, e.g., that true propositions tell it as it is or that they correspond with the facts. This gets us somewhere, perhaps, but “Truth is correspondence with fact” will remain a

platitude unless we can say what “correspondence” and “fact” mean in terms that don’t already presuppose an understanding of truth.

A moment’s reflection indicates how difficult that task is. One reason is that truth is an extremely basic concept. It is difficult to engage in any theoretical inquiry without employing it. You cannot even argue over a theory of truth without using the concept, because to question a theory is to question its truth, and to endorse a theory is to endorse it as true. In comparison, we can easily discuss what it is to be a person, or the nature of justice, without employing those concepts while doing so. But we cannot get behind the concept of truth as we can with these other concepts.

It seems that few concepts are as tightly wound into our thought as truth. Truth, for instance, is deeply connected to belief: when the witness tells us what she believes, this implies that she is reporting what she believes to be true. Similarly with assertion or endorsement: when we assert, we present ourselves as speaking the truth. Truth is also connected to knowledge: one doesn’t know that the butler did it unless it is really true that the butler did it. Truth is the central concept of logic as well: an argument is valid in the sense logicians are concerned with just when it is impossible for its premises to be true and its conclusion false. And as our platitudes about truth tell us, truth is related to that other mysterious concept, reality. To speak the truth is to speak of reality as it is. The fact that truth is so tightly interconnected with so many other philosophically interesting concepts is another reason why truth seems deep and why it seems important to understand what truth is.

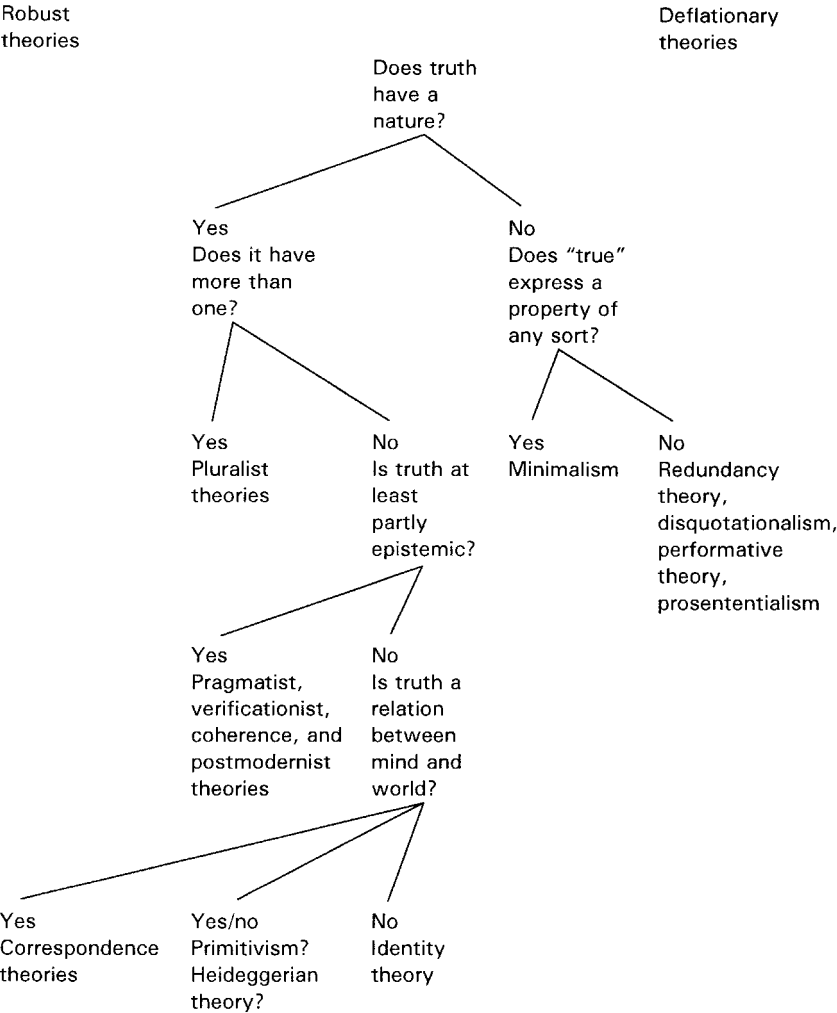
This connection between truth and other issues often muddies the very philosophical waters we are attempting to measure. Philosophers are frequently interested in different subjects when they ask about truth, subjects that involve the connection between truth and other areas of philosophical interest. Thus some philosophers who portray themselves as working on truth are actually interested in how we acquire truth, or in justification and knowledge; others are curious about the relation of truth to linguistic meaning, while still others wonder about the relation between truth and logic. These are all important issues, but none are the main focus of this book. For in each of the cases above, the issue is the

explanatory role of truth rather than its nature. When confronting those issues, we assume prior knowledge of what truth is.

But what does it mean to ask what truth is? In general, whenever we ask what something is, there are two questions we might be interested in. Suppose that I ask you what gold is. I might want to understand the *concept* of gold—what the word “gold” means in ordinary English. Alternatively, I might want to know about the underlying nature of the *property* of being gold—the substantive facts about gold, e.g., that it is an element with atomic number 79. Of course, these projects needn’t be completely distinct: my concept of gold presumably picks out many important and substantive facts about gold, e.g., that gold is a malleable yellow metal, for instance. Yet it also seems clear that I could have a good grasp on the concept of gold without knowing *all* the facts about its underlying nature.

When philosophers ask what truth is, they are interested sometimes in the concept, sometimes in the underlying nature of the property, and sometimes in both. In the case of gold, giving an analysis of the concept (for instance, by supplying necessary and sufficient conditions for the application of the word) needn’t tell you everything about what the property of being gold consists in. But in the case of truth, it is somewhat trickier to say how theories of the concept and theories of the property relate. Unlike the case of gold, we have no independent, empirical access to the property of truth itself except via that concept. Thus disputes over the property of truth are frequently (but not always) fought on conceptual ground, over how we might best define the concept of truth.<sup>1</sup> According to this latter method, we learn about the property of truth by learning about the concept. On the other hand, we might hold that as in the case of gold, learning about the concept can tell us much about the property without necessarily telling us everything about that property.

Whichever methodological stance we take, there are two central questions one might ask about the property, or underlying nature, of truth. First, does truth even have a nature, and second, if it does, what sort of nature does it have? These two questions are the focus of two very different types of debates (see figure 1).



**Figure 1**  
How various theories answer questions on the nature of truth, and where their answers place them on the robust-deflationary continuum

The more traditional of these two debates is the second, concerning what sort of property truth is. Theories that try to answer this question are often called *robust* theories of truth, since they assume that truth is an important property that requires a substantive and complex explanation. Those who engage in constructing such theories are motivated by questions like the following: Is there such a thing as absolute truth, or is all truth in some way or other subjective or relative? What sort of relationship, if any, do true propositions have to the world? Are all truths verifiable by sense experience? Could it turn out that even our best theories could be false? And so on. Broadly speaking, these questions all concern the objectivity of truth. Thus the key issue for robust theories of truth is *realism* (see the introduction to part I for a definition of this term in relation to truth).

While the realism debate continues to be of central importance, much contemporary work on truth has to do with the question of whether truth even has a nature to explain. This is the other main debate over truth that one finds in this volume. Since the beginning of the last century, *deflationists* have suspected that the so-called problem of truth was really a pseudoproblem. Driven by the seemingly intractable disputes over the nature of truth, as well as by a broadly empiricist epistemological attitude, deflationists hold that there is no single robust property shared by all the propositions we take as true. Consequently, our concept of truth should not be understood as expressing such a property but be seen as fulfilling some other function. Put somewhat differently, robust theorists argue that the various mysteries of truth require substantive metaphysical explanation, while deflationists believe that no such explanation is needed. In their view, the alleged mysteries should be not explained but explained away.

Although distinguishing the realism debate from the deflationary debate is helpful, we must be careful not to oversimplify. Many of the authors in this volume are engaged in both debates. Further, there is a growing consensus among some philosophers that neither traditional robust theories nor deflationary theories are right. If so, then we must find new ways to think about this old concept.



**Note**

1. Contributors to this volume differ on this point. Some (William Alston, chap. 3, and Michael Devitt, chap. 25) draw a clear line between these tasks; others do not. The latter is the more traditional tactic, e.g., of Russell (chap. 1). The basic idea is that by providing a generalized definition of the concept, one that says, roughly, that all and only true propositions have *F*, we could say that the nature of truth is *F*.

**Suggestions for Further Reading**

Blackburn, S., and Simmons, K., eds. 1999. *Truth*. Oxford: Oxford University Press. An excellent collection of papers concentrating on deflationism. The introduction by the editors is highly recommended.

Horwich, P., ed. 1994. *Theories of Truth*. Dartmouth Publishing Co. A difficult-to-find but comprehensive collection of classic papers.

Kirkham, R. 1992. *Theories of Truth*. Cambridge: MIT Press. An outstanding introduction to the various debates. Highly recommended.

Pitcher, G. 1964. *Truth*. Englewood Cliffs: Prentice-Hall. A classic collection of papers, sadly out of print. Many of these papers are included in the present volume.

Schmitt, F. 1995. *Truth: A Primer*. Boulder: Westview Press. A nicely concise introduction to some of the main debates.

I

---

## Realism and the Correspondence Theory



---

# Realism and the Correspondence Theory:

## Introduction

The root intuition behind *alethic realism* is that truth hinges not on us but on the world. A proposition is true in this sense when things in the world are as that proposition says they are. Some aspect of reality must simply be a certain way—if it is, then the proposition is true; if not, the proposition is false. Minimally speaking, this implies that truth has a nature and that its nature is objective: whether a proposition is true (in most cases) does not depend on what anyone believes.

The most venerable realist view is certainly the *correspondence theory of truth*, the view that a proposition is true just when it corresponds to reality. This is frequently said to be the point of Aristotle's claim, "To say that that which is, is not, or that which is not is, is a falsehood; and to say that that which is, is, and that which is not is not, is true."

Of course, Aristotle's remark and the thought "Truth is correspondence with reality" are no more than platitudes that few would deny. To flesh these intuitions out into a full-blown theory of truth, correspondence theorists must spell out three implicit metaphysical aspects of their position: they must say something about what has the property (the truth bearer), its correspondence (the truth relation), and the "reality" to which it corresponds (the truth maker).

In his classic essay, Bertrand Russell argues that it is beliefs that are true or false and facts that make beliefs true. Believing, for Russell, always consists in a believer's relation to two or more objects (what he sometimes confusingly calls the "terms" of the belief) united by another relation. Thus, my belief that *A* loves *B* consists in my being related to *A*, *B* and the relation *loving*. Objects related in this way form a "complex unity"; when the objects are related in the exact same order as they are in

my belief, then “this complex unity is called the fact corresponding to the belief.” A belief is true when it corresponds to a certain complex unity (a fact) and false when it does not.

Correspondence in Russell’s theory consists in a structural isomorphism or the *congruence* between the parts of a belief and the parts of a fact. True beliefs fit facts as a hand fits a glove. Russell, however, pays a high price for this structural fit between true beliefs and facts: for the objects of any belief and the objects that compose a complex unity are the same. How then do we understand my (false) belief that *A* loves *B* when there is no such person as *B*? If there is no *B*, it seems that on Russell’s theory, I can’t even have the belief that *A* loves *B*, since any such belief would involve a relation between myself and *B*. Furthermore, how exactly is it that *A*, *B*, and loving *combine* to form the particular “unity” or fact that *A* loves *B*, as opposed to the unity of *B*’s loving *A*? We might wonder what distinguishes these facts.

Unlike Russell, J. L. Austin takes correspondence to be a matter of *correlation* between whole statements, or “sentences as used by a certain person on a certain occasion” and whole facts, or “particular states of affairs.” Correspondence here is not structural: states of affairs and statements don’t “fit” one another because of the relations of their parts; they match up one to one. Further, Austin insists that correspondence is conventional. Our linguistic conventions determine whether the particular state of affairs I am referring to is of the appropriate type to make my statement true. In his response to Austin, P. F. Strawson (chap. 19) points out that it is not very clear what facts or states of affairs actually are. To talk of the fact or state of affairs that the cat is on the mat seems just like another way of talking about the true statement that the cat is on the mat. As a consequence, Strawson argues, facts sound suspiciously like mere linguistic shadows of true propositions.

An influential attempt to explain correspondence in a reductive, physicalist way can be found in Field’s article on Tarski (chap. 16). This is the theory of *correspondence as causal relation*. Motivated by a commitment to a global physicalist ontology, the view has two aspects. First, the concept of truth for natural languages is understood by way of a Tarskian recursive definition. Roughly speaking, the truth of a sentence consists in the *reference* of its parts to an objective reality. The second

aspect of the theory takes the reference relation itself to be a physical or causal relation between words and the world (see also Devitt, chap. 25). In short, the view is that a sentence token (a particular use of a sentence) correctly represents reality if and only if its component parts bear an appropriate causal relation to certain objects in the world.

William Alston takes a complete different tack in his essay. Alston argues that just as we distinguish between the concept or meaning of the word “gold” and the property of gold itself, so we should distinguish between the concept and property of truth. In Alston’s view, we should be no more surprised that our ordinary concept of truth fails to reveal the nature of correspondence or representation than we are that our ordinary concept of gold fails to reveal that gold in essence is the element with atomic number 79. Reflection on our ordinary concept of truth tells us only that propositions are true when the world is as they say it is, and that truth is a real property of propositions. Nonetheless, Alston argues forcefully that his minimal theory of truth qualifies as a type of realism. First, while our concept of truth does not give us specific information about the underlying nature of truth, the concept does imply that truth has a nature. Second, the concept provides a constraint on how that nature can be understood: truth is objective in the sense that whether a proposition is true does not depend on whether anyone is justified in believing the proposition. Thus on Alston’s account, as on the correspondence accounts discussed above, truth is mind-independent in the sense that, as Putnam says, it is “radically nonepistemic” (see chap. 11).

Most realists about truth *extend* this mind-independence to the truth makers as well. Facts, states of affairs, or propertyed objects are taken to be as they are independently of human concepts or beliefs. The concept of independence here is often cashed out counterfactually: “Stars are mind-independent” means that if there were no minds, there would still be stars. This view of truth makers, however, is best seen as a form of ontological realism: it is realism applied not to truth but to what exists in the world.

While ontological realism is often thought to entail a correspondence theory of truth and vice versa, it is worth noting that these views can be pried apart. Strictly speaking, one could believe that objects are partly or wholly constituted by their relations to concepts and still hold a corre-

spondence or realist account of truth. For suppose that objects and facts were so constituted. Then what true propositions correspond to would be mind-dependent. But that does not entail without further argument that whether or not propositions correspond to these mind-dependent entities depends on what we believe about those propositions. In this case, propositions would still be made true by, e.g., the obtaining of a certain state of affairs; that this state of affairs is mind-dependent is irrelevant (see Lynch 1998).

Realist views of truth face a number of objections. As we've seen, one can attack particular correspondence theories by criticizing their candidates for truth bearers, truth makers, or the correspondence relation. But correspondence theories also face difficult objections in virtue of their implicit commitment to realism about truth. Perhaps the oldest criticism of this type is that they entail a debilitating form of global skepticism. In its traditional form, the objection is that since we cannot step outside of our beliefs, we cannot ever check to see if they correspond to the world or not. Therefore, we can never know whether our beliefs are true. A more contemporary way of putting this problem is due to Putnam (chap. 11). According to realism about truth, the truth of every proposition is completely independent of the justification we may or may not have for believing that proposition. If so, then it is possible that any and all of our beliefs could be false—no matter how well justified they are. Some philosophers believe that radical skepticism of this sort is either impossible to answer or absurd, and they therefore are suspicious of any view that entails it. (Those who ascribe to coherence or pragmatist theories of truth, for instance, often cite this worry as the reason they reject realism. See parts II and III.<sup>1</sup>)

Another general problem for realist theories of truth concerns their scope. Traditional correspondence theories take correspondence to be the nature of truth for every proposition. But propositions, as we noted above, come in many varieties. Thus whatever correspondence consists in, it must be a property that can be had by propositions as diverse as *Two and two make four*, *Sherlock Holmes is smarter than Watson*, and *Sexism is wrong*. What sort of relation could possibly have relata that include abstract objects like numbers, fictional characters like Holmes, and a property like sexism? Note that this is a particularly difficult prob-

lem for those attracted to the causal/referential variety of correspondence. Presumably, what makes it true that two and two make four is a certain fact about the number two and the addition function. But numbers and functions are not physical objects, and hence are unlikely to enter into any causal relations with anything (Benacerraf 1973). Yet surely mathematical propositions can be true. Further still, it can seem that any realist theory will have problems accounting for the truths of economics, law, and morality. Yet these sorts of facts arguably concern entities that are in some sense mind-independent.

It is precisely this sort of problem that leads Terence Horgan in his essay to present an original version of the correspondence theory that allows for both direct and indirect word to world relations. In Horgan's view, which he calls "contextual semantics," truth in any discourse is determined jointly by the world and the semantic standards of the discourse. In short, truth is semantic correctness. Semantic correctness is a realist notion of truth, since it involves a type of correspondence with the world: it is never epistemic. Nonetheless, the type of correspondence can vary according to what we are talking about. This is because the semantic norms governing truth can vary with context. Thus, there is a spectrum of ways in which statements can correspond to the world. On one end of the spectrum are statements governed by maximally strict semantic standards. Such statements are true just when they directly correspond, via causal/referential relations, to mind-independent objects and properties. On the other end are statements whose truth is determined almost entirely by the semantic standards alone. In between sits the majority of the statements we make in life, such as those about corporations and works of art, which indirectly correspond to entities and attributes that are in many cases mind-dependent.

Horgan's view is therefore pluralist in a sense, and the reader is urged to compare it to the somewhat similar views of Wright (chap. 32), Lynch, (chap. 31), and Putnam (chap. 30). But there remains an important difference. Horgan's spectrum of correspondence is strongly hierarchical. Only statements that directly correspond to mind-independent entities accurately report how the world ultimately is. Thus, as Horgan notes, contextual semantics is best seen as a combination of metaphysical realism and a new version of the correspondence theory of truth.



## Note

1. A related problem for realist theories of truth is noted by Dummett (chap. 10). It is common to hold that understanding a statement consists in knowing its truth conditions. To know the truth conditions of a statement is to know the conditions under which that statement is either true or false. But if by “true” the realist means “true independently of what we justifiedly believe,” it seems mysterious how we could come to know the conditions under which many of our statements are true.

## Further Reading for Part I

Acton, H. B. 1934–1935. “The Correspondence Theory of Truth.” *Proceedings of the Aristotelian Society* 35: 177–194.

Alston, W. P. 1979. “Yes, Virginia, There Is a Real World.” *Proceedings of the American Philosophical Association* 52: 779–808.

Appiah, A. 1986. *For Truth in Semantics*. Oxford: Blackwell.

Armstrong, D. M. 1997. *A World of States of Affairs*. Cambridge: Cambridge University Press.

Austin, J. L. 1970. “Unfair to Facts.” In his *Philosophical Papers*, edited by J. O. Urmson and G. Warnock, pp. 154–174. Oxford: Oxford University Press.

Bealer, G. 1982. *Quality and Concept*. Oxford: Oxford University Press.

Benacerraf, P. 1973. “Mathematical Truth.” *Journal of Philosophy* 70: 661–679.

Cortens, A. 1999. *Global Anti-realism*. Boulder: Westview Press.

Davidson, D. 1984. “True to the Facts.” In his *Inquiries into Truth and Interpretation*. New York: Oxford University Press.

Devitt, M. 1997. *Realism and Truth*. 2nd ed. Princeton: Princeton University Press.

Dummett. 1978. *Truth and Other Enigmas*. London: Duckworth.

Field, H. 1974. “Quine and the Correspondence Theory.” *Philosophical Review* 83: 200–228.

Forbes, G. 1986. “Truth, Correspondence, and Redundancy.” In *Fact, Science, and Morality*, edited by C. Wright and G. MacDonald. Oxford: Blackwell.

Horgan, T. 1986. “Truth and Ontology.” *Philosophical Papers* 15: 1–21.

Horgan, T. 1991. “Metaphysical Realism and Psychologistic Semantics.” *Erkenntnis* 34: 297–322.

Lynch, M. 1997. “Realistic Minimalism or Minimal Realism?” *Philosophical Quarterly* 47: 512–518.

Lynch, M. 1998. *Truth in Context*. Cambridge: MIT Press.

Moore, G. E. 1953. Chapter 3 of *Some Main Problems in Philosophy*. New York: Macmillan Co.

- Pendlebury, M. 1986. "Facts as Truthmakers." *Monist* 69: 177–188.
- Putnam, H. 1981. *Realism, Truth, and History*. Cambridge: Cambridge University Press.
- Sosa, E. 1993. "Epistemology, Realism, and Truth." In *Language and Logic*, edited by James E. Tomberlin. Atascadero, Calif.: Ridgeview Publishing Co.
- Russell, Bertrand. 1906–1907. "On the Nature of Truth." *Proceedings of the Aristotelian Society* 7: 28–49.
- Russell, B. 1948. *Inquiry into Meaning and Truth*.
- Sellars, W. 1963. "Truth and 'Correspondence'." In his *Science, Perception, and Reality*. New York: Humanities Press.
- Wiggins, D. "What Would Be a Substantial Theory of Truth?" In Van Stratten, ed, *Philosophical Subjects*, pp. 189–221. Oxford: Oxford University Press.
- Van Inwagen, P. 1988. "On Always Being Wrong." *Midwest Studies in Philosophy* 12: 95–112.
- Wittgenstein, L. 1922. *Tractatus Logico-philosophicus*. London: Routledge and Kegan Paul.



---

# Truth and Falsehood

Bertrand Russell

Our knowledge of truths, unlike our knowledge of things, has an opposite, namely *error*. So far as things are concerned, we may know them or not know them, but there is no positive state of mind which can be described as erroneous knowledge of things, so long, at any rate, as we confine ourselves to knowledge by acquaintance. Whatever we are acquainted with must be something; we may draw wrong inferences from our acquaintance, but the acquaintance itself cannot be deceptive. Thus there is no dualism as regards acquaintance. But as regards knowledge of truths, there is a dualism. We may believe what is false as well as what is true. We know that on very many subjects different people hold different and incompatible opinions: hence some beliefs must be erroneous. Since erroneous beliefs are often held just as strongly as true beliefs, it becomes a difficult question how they are to be distinguished from true beliefs. How are we to know, in a given case, that our belief is not erroneous? This is a question of the very greatest difficulty, to which no completely satisfactory answer is possible. There is, however, a preliminary question which is rather less difficult, and that is: What do we *mean* by truth and falsehood? It is this preliminary question which is to be considered in this chapter.

In this chapter we are not asking how we can know whether a belief is true or false: we are asking what is meant by the question whether a belief is true or false. It is to be hoped that a clear answer to this question may help us to obtain an answer to the question what beliefs are true, but for the present we ask only 'What is truth?' and 'What is falsehood?' not 'What beliefs are true?' and 'What beliefs are false?' It is very important to keep these different questions entirely separate, since any confusion

between them is sure to produce an answer which is not really applicable to either.

There are three points to observe in the attempt to discover the nature of truth, three requisites which any theory must fulfil.

1. Our theory of truth must be such as to admit of its opposite, falsehood. A good many philosophers have failed adequately to satisfy this condition: they have constructed theories according to which all our thinking ought to have been true, and have then had the greatest difficulty in finding a place for falsehood. In this respect our theory of belief must differ from our theory of acquaintance, since in the case of acquaintance it was not necessary to take account of any opposite.

2. It seems fairly evident that if there were no beliefs there could be no falsehood, and no truth either, in the sense in which truth is correlative to falsehood. If we imagine a world of mere matter, there would be no room for falsehood in such a world, and although it would contain what may be called 'facts', it would not contain any truths, in the sense in which truths are things of the same kind as falsehoods. In fact, truth and falsehood are properties of beliefs and statements: hence a world of mere matter, since it would contain no beliefs or statements, would also contain no truth or falsehood.

3. But, as against what we have just said, it is to be observed that the truth or falsehood of a belief always depends upon something which lies outside the belief itself. If I believe that Charles I died on the scaffold, I believe truly, not because of any intrinsic quality of my belief, which could be discovered by merely examining the belief, but because of an historical event which happened two and a half centuries ago. If I believe that Charles I died in his bed, I believe falsely: no degree of vividness in my belief, or of care in arriving at it, prevents it from being false, again because of what happened long ago, and not because of any intrinsic property of my belief. Hence, although truth and falsehood are properties of beliefs, they are properties dependent upon the relations of the beliefs to other things, not upon any internal quality of the beliefs.

The third of the above requisites leads us to adopt the view—which has on the whole been commonest among philosophers—that truth consists in some form of correspondence between belief and fact. It is, however, by no means an easy matter to discover a form of correspondence to which there are no irrefutable objections. By this partly—and partly by the feeling that, if truth consists in a correspondence of thought with something outside thought, thought can never know when truth has

been attained—many philosophers have been led to try to find some definition of truth which shall not consist in relation to something wholly outside belief. The most important attempt at a definition of this sort is the theory that truth consists in *coherence*. It is said that the mark of falsehood is failure to cohere in the body of our beliefs, and that it is the essence of a truth to form part of the completely rounded system which is The Truth.

There is, however, a great difficulty in this view, or rather two great difficulties. The first is that there is no reason to suppose that only *one* coherent body of beliefs is possible. It may be that, with sufficient imagination, a novelist might invent a past for the world that would perfectly fit on to what we know, and yet be quite different from the real past. In more scientific matters, it is certain that there are often two or more hypotheses which account for all the known facts on some subject, and although, in such cases, men of science endeavour to find facts which will rule out all the hypotheses except one, there is no reason why they should always succeed.

In philosophy, again, it seems not uncommon for two rival hypotheses to be both able to account for all the facts. Thus, for example, it is possible that life is one long dream, and that the outer world has only that degree of reality that the objects of dreams have; but although such a view does not seem inconsistent with known facts, there is no reason to prefer it to the common-sense view, according to which other people and things do really exist. Thus coherence as the definition of truth fails because there is no proof that there can be only one coherent system.

The other objection to this definition of truth is that it assumes the meaning of 'coherence' known, whereas, in fact, 'coherence' presupposes the truth of the laws of logic. Two propositions are coherent when both may be true, and are incoherent when one at least must be false. Now in order to know whether two propositions can both be true, we must know such truths as the law of contradiction. For example, the two propositions, 'this tree is a beech' and 'this tree is not a beech', are not coherent, because of the law of contradiction. But if the law of contradiction itself were subjected to the test of coherence, we should find that, if we choose to suppose it false, nothing will any longer be incoherent with anything else. Thus the laws of logic supply the skeleton or frame-

work within which the test of coherence applies, and they themselves cannot be established by this test.

For the above two reasons, coherence cannot be accepted as giving the *meaning* of truth, though it is often a most important *test* of truth after a certain amount of truth has become known.

Hence we are driven back to *correspondence with fact* as constituting the nature of truth. It remains to define precisely what we mean by 'fact', and what is the nature of the correspondence which must subsist between belief and fact, in order that belief may be true.

In accordance with our three requisites, we have to seek a theory of truth which (1) allows truth to have an opposite, namely falsehood, (2) makes truth a property of beliefs, but (3) makes it a property wholly dependent upon the relation of the beliefs to outside things.

The necessity of allowing for falsehood makes it impossible to regard belief as a relation of the mind to a single object, which could be said to be what is believed. If belief were so regarded, we should find that, like acquaintance, it would not admit of the opposition of truth and falsehood, but would have to be always true. This may be made clear by examples. Othello believes falsely that Desdemona loves Cassio. We cannot say that this belief consists in a relation to a single object, 'Desdemona's love for Cassio', for if there were such an object, the belief would be true. There is in fact no such object, and therefore Othello cannot have any relation to such an object. Hence his belief cannot possibly consist in a relation to this object.

It might be said that his belief is a relation to a different object, namely 'that Desdemona loves Cassio'; but it is almost as difficult to suppose that there is such an object as this, when Desdemona does not love Cassio, as it was to suppose that there is 'Desdemona's love for Cassio'. Hence it will be better to seek for a theory of belief which does not make it consist in a relation of the mind to a single object.

It is common to think of relations as though they always held between *two* terms, but in fact this is not always the case. Some relations demand three terms, some four, and so on. Take, for instance, the relation 'between'. So long as only two terms come in, the relation 'between' is impossible: three terms are the smallest number that render it possible. York is between London and Edinburgh; but if London and Edinburgh

were the only places in the world, there could be nothing which was between one place and another. Similarly *jealousy* requires three people: there can be no such relation that does not involve three at least. Such a proposition as 'A wishes B to promote C's marriage with D' involves a relation of four terms; that is to say, A and B and C and D all come in, and the relation involved cannot be expressed otherwise than in a form involving all four. Instances might be multiplied indefinitely, but enough has been said to show that there are relations which require more than two terms before they can occur.

The relation involved in *judging* or *believing* must, if falsehood is to be duly allowed for, be taken to be a relation between several terms, not between two. When Othello believes that Desdemona loves Cassio, he must not have before his mind a single object, 'Desdemona's love for Cassio', or 'that Desdemona loves Cassio', for that would require that there should be objective falsehoods, which subsist independently of any minds; and this, though not logically refutable, is a theory to be avoided if possible. Thus it is easier to account for falsehood if we take judgement to be a relation in which the mind and the various objects concerned all occur severally; that is to say, Desdemona and loving and Cassio must all be terms in the relation which subsists when Othello believes that Desdemona loves Cassio. This relation, therefore, is a relation of four terms, since Othello also is one of the terms of the relation. When we say that it is a relation of four terms, we do not mean that Othello has a certain relation to Desdemona, and has the same relation to loving and also to Cassio. This may be true of some other relation than believing; but believing, plainly, is not a relation which Othello has to *each* of the three terms concerned, but to *all* of them together: there is only one example of the relation of believing involved, but this one example knits together four terms. Thus the actual occurrence, at the moment when Othello is entertaining his belief, is that the relation called 'believing' is knitting together into one complex whole the four terms Othello, Desdemona, loving, and Cassio. What is called belief or judgement, is nothing but this relation of believing or judging, which relates a mind to several things other than itself. An *act* of belief or of judgement is the occurrence between certain terms at some particular time, of the relation of believing or judging.



We are now in a position to understand what it is that distinguishes a true judgement from a false one. For this purpose we will adopt certain definitions. In every act of judgement there is a mind which judges, and there are terms concerning which it judges. We will call the mind the *subject* in the judgement, and the remaining terms the *objects*. Thus, when Othello judges that Desdemona loves Cassio, Othello is the subject, while the objects are Desdemona and loving and Cassio. The subject and the objects together are called the *constituents* of the judgement. It will be observed that the relation of judging has what is called a 'sense' or 'direction'. We may say, metaphorically, that it puts its objects in a certain *order*, which we may indicate by means of the order of the words in the sentence. (In an inflected language, the same thing will be indicated by inflections, e.g. by the difference between nominative and accusative.) Othello's judgement that Cassio loves Desdemona differs from his judgement that Desdemona loves Cassio, in spite of the fact that it consists of the same constituents, because the relation of judging places the constituents in a different order in the two cases. Similarly, if Cassio judges that Desdemona loves Othello, the constituents of the judgement are still the same, but their order is different. This property of having a 'sense' or 'direction' is one which the relation of judging shares with all other relations. The 'sense' of relations is the ultimate source of order and series and a host of mathematical concepts; but we need not concern ourselves further with this aspect.

We spoke of the relation called 'judging' or 'believing' as knitting together into one complex whole the subject and the objects. In this respect, judging is exactly like every other relation. Whenever a relation holds between two or more terms, it unites the terms into a complex whole. If Othello loves Desdemona, there is such a complex whole as 'Othello's love for Desdemona'. The terms united by the relation may be themselves complex, or may be simple, but the whole which results from their being united must be complex. Wherever there is a relation which relates certain terms, there is a complex object formed of the union of those terms; and conversely, wherever there is a complex object, there is a relation which relates its constituents. When an act of believing occurs, there is a complex, in which 'believing' is the uniting relation, and subject

and objects are arranged in a certain order by the 'sense' of the relation of believing. Among the objects, as we saw in considering 'Othello believes that Desdemona loves Cassio', one must be a relation—in this instance, the relation 'loving'. But this relation, as it occurs in the act of believing, is not the relation which creates the unity of the complex whole consisting of the subject and the objects. The relation 'loving', as it occurs in the act of believing, is one of the objects—it is a brick in the structure, not the cement. The cement is the relation 'believing'. When the belief is *true*, there is another complex unity, in which the relation which was one of the objects of the belief relates the other objects. Thus, e.g., if Othello believes *truly* that Desdemona loves Cassio, then there is a complex unity, 'Desdemona's love for Cassio', which is composed exclusively of the *objects* of the belief, in the same order as they had in the belief, with the relation which was one of the objects occurring now as the cement that binds together the other objects of the belief. On the other hand, when a belief is *false*, there is no such complex unity composed only of the objects of the belief. If Othello believes *falsely* that Desdemona loves Cassio, then there is no such complex unity as 'Desdemona's love for Cassio'.

Thus a belief is *true* when it *corresponds* to a certain associated complex, and *false* when it does not. Assuming, for the sake of definiteness, that the objects of the belief are two terms and a relation, the terms being put in a certain order by the 'sense' of the believing, then if the two terms in that order are united by the relation into a complex, the belief is true; if not, it is false. This constitutes the definition of truth and falsehood that we were in search of. Judging or believing is a certain complex unity of which a mind is a constituent; if the remaining constituents, taken in the order which they have in the belief, form a complex unity, then the belief is true; if not, it is false.

Thus although truth and falsehood are properties of beliefs, yet they are in a sense extrinsic properties, for the condition of the truth of a belief is something not involving beliefs, or (in general) any mind at all, but only the *objects* of the belief. A mind, which believes, believes truly when there is a *corresponding* complex not involving the mind, but only its objects. This correspondence ensures truth, and its absence entails false-

hood. Hence we account simultaneously for the two facts that beliefs (a) depend on minds for their *existence*, (b) do not depend on minds for their, *truth*.

We may restate our theory as follows: If we take such a belief as 'Othello believes that Desdemona loves Cassio', we will call Desdemona and Cassio the *object-terms*, and loving the *object-relation*. If there is a complex unity 'Desdemona's love for Cassio', consisting of the object-terms related by the object-relation in the same order as they have in the belief, then this complex unity is called the *fact corresponding to the belief*. Thus a belief is true when there is a corresponding fact, and is false when there is no corresponding fact.

It will be seen that minds do not *create* truth or falsehood. They create beliefs, but when once the beliefs are created, the mind cannot make them true or false, except in the special case where they concern future things which are within the power of the person believing, such as catching trains. What makes a belief true is a *fact*, and this fact does not (except in exceptional cases) in any way involve the mind of the person who has the belief.

Having now decided what we *mean* by truth and falsehood, we have next to consider what ways there are of knowing whether this or that belief is true or false. This consideration will occupy the next chapter.

---

# Truth

J. L. Austin

## 1

‘What is truth?’ said jesting Pilate, and would not stay for an answer. Pilate was in advance of his time. For ‘truth’ itself is an abstract noun, a camel, that is, of a logical construction, which cannot get past the eye even of a grammarian. We approach it cap and categories in hand: we ask ourselves whether Truth is a substance (the Truth, the Body of Knowledge), or a quality (something like the color red, inhering in truths), or a relation (‘correspondence’).<sup>1</sup> But philosophers should take something more nearly their own size to strain at. What needs discussing rather is the use, or certain uses, of the word ‘true.’ *In vino*, possibly, ‘*veritas*,’ but in a sober symposium ‘*verum*.’

## 2

What is it that we say is true or is false? Or, how does the phrase ‘is true’ occur in English sentences? The answers appear at first multifarious. We say (or are said to say) that beliefs are true, that descriptions or accounts are true, that propositions or assertions or statements are true, and that words or sentences are true: and this is to mention only a selection of the more obvious candidates. Again, we say (or are said to say) ‘It is true that the cat is on the mat,’ or ‘It is true to say that the cat is on the mat,’ or “‘The cat is on the mat’ is true.’ We also remark on occasion, when someone else has said something, ‘Very true’ or ‘That’s true’ or ‘True enough.’

Most (though not all) of these expressions, and others besides, certainly do occur naturally enough. But it seems reasonable to ask whether there is not some use of 'is true' that is primary, or some generic name for that which at bottom we are always saying 'is true.' Which, if any, of these expressions is to be taken *au pied de la lettre*? To answer this will not take us long, nor, perhaps, far: but in philosophy the foot of the letter is the foot of the ladder.

I suggest that the following are the primary forms of expression:

It is true (to say) that the cat is on the mat.

That statement (of his, etc.) is true.

The statement that the cat is on the mat is true.

But first for the rival candidates.

a. Some say that 'truth is primarily a property of beliefs.' But it may be doubted whether the expression 'a true belief' is at all common outside philosophy and theology: and it seems clear that a man is said to hold a true belief when and in the sense that he believes (in) *something which* is true, or believes that *something which* is true is true. Moreover if, as some also say, a belief is 'of the nature of a picture,' then it is of the nature of what cannot be true, though it may be, for example, faithful.<sup>2</sup>

b. True descriptions and true accounts are simply varieties of true statements or of collections of true statements, as are true answers and the like. The same applies to propositions too, in so far as they are genuinely said to be true (and not, as more commonly, sound, tenable and so on).<sup>3</sup> A proposition in law or in geometry is something portentous, usually a generalization, that we are invited to accept and that has to be recommended by argument: it cannot be a direct report on current observation—if you look and inform me that the cat is on the mat, that is not a proposition though it is a statement. In philosophy, indeed, 'proposition' is sometimes used in a special way for 'the meaning or sense of a sentence or family of sentences': but whether we think a lot or little of this usage, a proposition in this sense cannot, at any rate, be what we say is true or false. For we never say 'The meaning (or sense) of this sentence (or of these words) is true': what we do say is what the judge or jury says, namely that '*The words* taken in this sense, or if we assign to them such and such a meaning, or so interpreted or understood, *are true*.'

c. Words and sentences are indeed said to be true, the former often, the latter rarely. But only in certain senses. Words as discussed by philologists, or by lexicographers, grammarians, linguists, phoneticians, printers,

critics (stylistic or textual) and so on, are not true or false: they are wrongly formed, or ambiguous or defective or untranslatable or unprouncable or misspelled or archaistic or corrupt or what not.<sup>4</sup> Sentences in similar contexts are elliptic or involved or alliterative or ungrammatical. We may, however, genuinely say 'His closing words were very true' or 'The third sentence on page 5 of his speech is quite false': but here 'words' and 'sentence' refer, as is shown by the demonstratives (possessive pronouns, temporal verbs, definite descriptions, etc.), which in this usage consistently accompany them, to the words or sentence *as used by a certain person on a certain occasion*. That is, they refer (as does 'Many a true word spoken in jest') to *statements*.

A statement is made and its making is an historic event, the utterance by a certain speaker or writer of certain words (a sentence) to an audience with reference to an historic situation, event or what not.<sup>5</sup>

A sentence is made *up of* words, a statement is made *in* words. A sentence is not English or not good English, a statement is not in English or not in good English. Statements are made, words or sentences are used. We talk of *my* statement, but of *the English* sentence (if a sentence is mine, I coined it, but I do not coin statements). The *same* sentence is used in making *different* statements (I say 'It is mine,' you say 'It is mine'): it may also be used on two occasions or by two persons in making the *same* statement, but for this the utterance must be made with reference to the same situation or event.<sup>6</sup> We speak of 'the statement that S,' but of 'the sentence "S,"' not of 'the sentence that S.'<sup>7</sup>

When I say that a statement is what is true, I have no wish to become wedded to one word. 'Assertion,' for example, will in most contexts do just as well, though perhaps it is slightly wider. Both words share the weakness of being rather solemn (much more so than the more general 'what you said' or 'your words')—though perhaps we are generally being a little solemn when we discuss the truth of anything. Both have the merit of clearly referring to the historic use of a sentence by an utterer, and of being therefore precisely not equivalent to 'sentence.' For it is a fashionable mistake to take as primary '(The sentence) "S" is true (in the English language).' Here the addition of the words 'in the English language' serves to emphasize that 'sentence' is not being used as equivalent to 'statement,' so that it precisely is not what can be true or false (and moreover, 'true in the English language' is a solecism, mismodeled presumably, and with deplorable effect, on expressions like 'true in geometry').

## 3

When is a statement true? The temptation is to answer (at least if we confine ourselves to 'straightforward' statements): 'When it corresponds to the facts.' And as a piece of standard English this can hardly be wrong. Indeed, I must confess I do not really think it is wrong at all: the theory of truth is a series of truisms. Still, it can at least be misleading.

If there is to be communication of the sort that we achieve by language at all, there must be a stock of symbols of some kind which a communicator ('the speaker') can produce 'at will' and which a communicatee ('the audience') can observe: these may be called the 'words,' though, of course, they need not be anything very like what we should normally call words—they might be signal flags, etc. There must also be something other than the words, which the words are to be used to communicate about: this may be called the 'world.' There is no reason why the world should not include the words, in every sense except the sense of the actual statement itself which on any particular occasion is being made about the world. Further, the world must exhibit (we must observe) similarities and dissimilarities (there could not be the one without the other): if everything were either absolutely indistinguishable from anything else or completely unlike anything else, there would be nothing to say. And finally (for present purposes—of course there are other conditions to be satisfied too) there must be two sets of conventions:

*Descriptive* conventions correlating the words (= sentences) with the *types* of situation, thing, event, etc., to be found in the world.

*Demonstrative* conventions correlating the words (= statements) with the *historic* situations, etc., to be found in the world.<sup>8</sup>

A statement is said to be true when the historic state of affairs to which it is correlated by the demonstrative conventions (the one to which it 'refers') is of a type<sup>9</sup> with which the sentence used in making it is correlated by the descriptive conventions.<sup>10</sup>

## 3a

Troubles arise from the use of the word 'facts' for the historic situations, events, etc., and in general, for the world. For 'fact' is regularly used in

conjunction with 'that' in the sentences 'The fact is that S' or 'It is a fact that S' and in the expression 'the fact that S,' all of which imply that it would be true to say that S.<sup>11</sup>

This may lead us to suppose that

- i. 'fact' is only an alternative expression for 'true statement.' We note that when a detective says 'Let's look at the facts' he does not crawl round the carpet, but proceeds to utter a string of statements: we even talk of 'stating the facts';
- ii. for every true statement there exists 'one' and its own precisely corresponding fact—for every cap the head it fits.

It is (i) which leads to some of the mistakes in 'coherence' or formalist theories; (ii) to some of those in 'correspondence' theories. Either we suppose that there is nothing there but the true statement itself, nothing to which it corresponds, or else we populate the world with linguistic *Doppelgänger* (and grossly overpopulate it—every nugget of 'positive' fact overlaid by a massive concentration of 'negative' facts, every tiny detailed fact larded with generous general facts, and so on).

When a statement is true, there is, *of course*, a state of affairs which makes it true and which is *toto mundo* distinct from the true statement about it: but equally of course, we can only *describe* that state of affairs *in words* (either the same or, with luck, others). I can only describe the situation in which it is true to say that I am feeling sick by saying that it is one in which I am feeling sick (or experiencing sensations of nausea):<sup>12</sup> yet between stating, however truly, that I am feeling sick and feeling sick there is a great gulf fixed.<sup>13</sup>

'Fact that' is a phrase designed for use in situations where the distinction between a true statement and the state of affairs about which it is a truth is neglected; as it often is with advantage in ordinary life, though seldom in philosophy—above all in discussing truth, where it is precisely our business to prize the words off the world and keep them off it. To ask 'Is the fact that S the true statement that S or that which it is true of?' may beget absurd answers. To take an analogy: although we may sensibly ask 'Do we *ride* the word "elephant" or the animal?' and equally sensibly 'Do we *write* the word or the animal?' it is nonsense to ask 'Do we *define* the word or the animal?' For defining an elephant (supposing we ever do this) is a compendious description of an operation involving



both word and animal (do we focus the image or the battleship?); and so speaking about 'the fact that' is a compendious way of speaking about a situation involving both words and world.<sup>14</sup>

### 3b

'Corresponds' also gives trouble, because it is commonly given too restricted or too colorful a meaning, or one which in this context it cannot bear. The only essential point is this: that the correlation between the words (= sentences) and the type of situation, event, etc. which is to be such that when a statement in those words is made with reference to an historic situation of that type the statement is then true, is *absolutely and purely* conventional. We are absolutely free to appoint *any* symbol to describe *any* type of situation, so far as merely being true goes. In a small one-spade language *tst nuts* might be true in exactly the same circumstances as the statement in English that the National Liberals are the people's choice.<sup>15</sup> There is no need whatsoever for the words used in making a true statement to 'mirror' in any way, however indirect, any feature whatsoever of the situation or event; a statement no more needs, in order to be true, to reproduce the 'multiplicity,' say, or the 'structure' or 'form' of the reality, than a word needs to be echoic or writing pictographic. To suppose that it does, is to fall once again into the error of reading back into the world the features of language.

The more rudimentary a language, the more, very often, it will tend to have a 'single' word for a highly 'complex' type of situation: this has such disadvantages as that the language becomes elaborate to learn and is incapable of dealing with situations which are nonstandard, unforeseen, for which there may just be no word. When we go abroad equipped only with a phrase-book, we may spend long hours learning by heart—

A<sup>1</sup>-moest-fa<sup>1</sup>nd-<sup>°</sup>tschâ<sup>r</sup>woum<sup>°</sup>n,  
Ma<sup>1</sup>hwîl-iz-wau<sup>r</sup>pt (bènt),

and so on and so on, yet faced with the situation where we have the pen of our aunt, find ourselves quite unable to say so. The characteristics of a more developed language (articulation, morphology, syntax, abstractions, etc.), do not make statements in it any more capable of being true

or capable of being any more true, they make it more adaptable, more learnable, more comprehensive, more precise, and so on; and *these* aims may no doubt be furthered by making the language (allowance made for the nature of the medium) 'mirror' in conventional ways features descried in the world.

Yet even when a language does 'mirror' such features very closely (and does it ever?) the truth of statements remains still a matter, as it was with the most rudimentary languages, of the words used being the ones *conventionally appointed* for situations of the type to which that referred to belongs. A picture, a copy, a replica, a photograph—these are *never* true in so far as they are reproductions, produced by natural or mechanical means: a reproduction can be accurate or lifelike (true *to* the original), as a gramophone recording or a transcription may be, but not true (*of*) as a record of proceedings can be. In the same way a (natural) sign *of* something can be infallible or unreliable but only an (artificial) sign *for* something can be right or wrong.<sup>16</sup>

There are many intermediate cases between a true account and a faithful picture, as here somewhat forcibly contrasted, and it is from the study of these (a lengthy matter) that we can get the clearest insight into the contrast. For example, maps: these may be called pictures, yet they are highly conventionalized pictures. If a map can be clear or accurate or misleading, like a statement, why can it not be true or exaggerated? How do the 'symbols' used in mapmaking differ from those used in statementmaking? On the other hand, if an air-mosaic is not a map, why is it not? And when does a map become a diagram? These are the really illuminating questions.

#### 4

Some have said that—

To say that an assertion is true is not to make any further assertion at all.

In all sentences of the form '*p* is true' the phrase 'is true' is logically superfluous.

To say that a proposition is true is just to assert it, and to say that it is false is just to assert its contradictory.

But wrongly. TstS (except in paradoxical cases of forced and dubious manufacture) refers to the world or any part of it exclusive of tstS, i.e., of itself.<sup>17</sup> TstST refers to the world or any part of it *inclusive* of tstS, though once again exclusive of itself, i.e., of tstST. That is, tstST refers to something to which tstS cannot refer. TstST does not, certainly, include any statement referring to the world exclusive of tstS which is not included already in tstS—more, it seems doubtful whether it does include that statement about the world exclusive of tstS which is made when we state that S. (If I state that tstS is true, should we really agree that I have stated that S? Only ‘by implication.’)<sup>18</sup> But all this does not go any way to show that tstST is not a statement different from tstS. If Mr. Q writes on a notice-board ‘Mr. W is a burglar,’ then a trial is held to decide whether Mr. Q’s published statement that Mr. W is a burglar is a libel: finding ‘Mr. Q’s statement was true (in substance and in fact).’ Thereupon a second trial is held, to decide whether Mr. W is a burglar, in which Mr. Q’s statement is no longer under consideration: verdict ‘Mr. W is a burglar.’ It is an arduous business to hold a second trial: why is it done if the verdict is the same as the previous finding?<sup>19</sup>

What is felt is that the evidence considered in arriving at the one verdict is the same as that considered in arriving at the other. This is not strictly correct. It is more nearly correct that whenever tstS is true then tstST is also true and conversely, and that whenever tstS is false tstST is also false and conversely.<sup>20</sup> And it is argued that the words ‘is true’ are logically superfluous because it is believed that generally if any two statements are always true together and always false together then they must mean the same. Now whether this is in general a sound view may be doubted: but even if it is, why should it not break down in the case of so obviously ‘peculiar’ a phrase as ‘is true’? Mistakes in philosophy notoriously arise through thinking that what holds of ‘ordinary’ words like ‘red’ or ‘grows’ must also hold of extraordinary words like ‘real’ or ‘exists.’ But that ‘true’ is just such another extraordinary word is obvious.<sup>21</sup>

There is something peculiar about the ‘fact’ which is described by tstST, something which may make us hesitate to call it a ‘fact’ at all; namely, that the relation between tstS and the world which tstST asserts to obtain is a *purely conventional* relation (one which ‘thinking makes

so'). For we are aware that this relation is one which we could alter at will, whereas we like to restrict the word 'fact' to *hard* facts, facts which are natural and unalterable, or anyhow not alterable at will. Thus, to take an analogous case, we may not like calling it a fact that the word elephant means what it does, though we can be induced to call it a (soft) fact—and though, of course, we have no hesitation in calling it a fact that contemporary English speakers use the word as they do.

An important point about this view is that it confuses falsity with negation: for according to it, it is the same thing to say 'He is not at home' as to say 'It is false that he is at home.' (But what if no one has said that he *is* at home? What if he is lying upstairs dead?) Too many philosophers maintain, when anxious to explain away negation, that a negation is just a second order affirmation (to the effect that a certain first order affirmation is false), yet, when anxious to explain away falsity, maintain that to assert that a statement is false is just to assert its negation (contradictory). It is impossible to deal with so fundamental a matter here.<sup>22</sup> Let me assert the following merely. Affirmation and negation are exactly on a level, in this sense, that no language can exist which does not contain conventions for both and that both refer to the world equally directly, not to statements about the world: whereas a language can quite well exist without any device to do the work of 'true' and 'false.' Any satisfactory theory of truth must be able to cope equally with falsity:<sup>23</sup> but 'is false' can only be maintained to be logically superfluous by making this fundamental confusion.

## 5

There is another way of coming to see that the phrase 'is true' is not logically superfluous, and to appreciate what sort of a statement it is to say that a certain statement is true. There are numerous other adjectives which are in the same class as 'true' and 'false,' which are concerned, that is, with the relations between the words (as uttered with reference to an historic situation) and the world, and which nevertheless no one would dismiss as logically superfluous. We say, for example, that a certain statement is exaggerated or vague or bald, a description somewhat rough or misleading or not very good, an account rather general or too concise.

In cases like these it is pointless to insist on deciding in simple terms whether the statement is 'true or false.' Is it true or false that Belfast is north of London? That the galaxy is the shape of a fried egg? That Beethoven was a drunkard? That Wellington won the battle of Waterloo? There are various *degrees and dimensions* of success in making statements: the statements fit the facts always more or less loosely, in different ways on different occasions for different intents and purposes. What may score full marks in a general knowledge test may in other circumstances get a gamma. And even the most adroit of languages may fail to 'work' in an abnormal situation or to cope, or cope reasonably simply, with novel discoveries: is it true or false that the dog goes round the cow?<sup>24</sup> What, moreover, of the large class of cases where a statement is not so much false (or true) as out of place, *inept* ('All the signs of bread' said when the bread is before us)?

We become obsessed with 'truth' when discussing statements, just as we become obsessed with 'freedom' when discussing conduct. So long as we think that what has always and alone to be decided is whether a certain action was done freely or was not, we get nowhere: but so soon as we turn instead to the numerous other adverbs used in the same connection ('accidentally,' 'unwillingly,' 'inadvertently,' etc.), things become easier, and we come to see that no concluding inference of the form 'Ergo, it was done freely (or not freely)' is required. Like freedom, truth is a bare minimum or an illusory ideal (the truth, the whole truth and nothing but the truth about, say, the battle of Waterloo or the *Primavera*).<sup>25</sup>

## 6

Not merely is it jejune to suppose that all a statement aims to be is 'true,' but it may further be questioned whether every 'statement' does aim to be true at all. The principle of Logic, that 'Every proposition must be true or false,' has too long operated as the simplest, most persuasive and most pervasive form of the descriptive fallacy. Philosophers under its influence have forcibly interpreted all 'propositions' on the model of the statement that a certain thing is red, as made when the thing concerned is currently under observation.

Recently, it has come to be realized that many utterances which have been taken to be statements (merely because they are not, on grounds of grammatical form, to be classed as commands, questions, etc.) are not in fact descriptive, nor susceptible of being true or false. When is a statement not a statement? When it is a formula in a calculus: when it is a performatory utterance: when it is a value-judgment: when it is a definition: when it is part of a work of fiction—there are many such suggested answers. It is simply not the business of such utterances to ‘correspond to the facts’ (and even genuine statements have other businesses besides that of so corresponding).

It is a matter for decision how far we should continue to call such masqueraders ‘statements’ at all, and how widely we should be prepared to extend the uses of ‘true’ and ‘false’ in ‘different senses.’ My own feeling is that it is better, when once a masquerader has been unmasked, *not* to call it a statement and *not* to say it is true or false. In ordinary life we should not call most of them statements at all, though philosophers and grammarians may have come to do so (or rather, have lumped them all together under the term of art ‘proposition’). We make a difference between ‘You said you promised’ and ‘You stated that you promised’: the former can mean that you said ‘I promise,’ whereas the latter must mean that you said ‘I promised’: the latter, which we say you ‘stated,’ is something which is true or false, whereas for the former, which is not true or false, we use the wider verb to ‘say.’ Similarly, there is a difference between ‘You say this is (call this) a good picture’ and ‘You state that this is a good picture.’ Moreover, it was only so long as the real nature of arithmetical formulas, say, or of geometrical axioms remained unrecognized, and they were thought to record information about the world, that it was reasonable to call them ‘true’ (and perhaps even ‘statements’—though were they ever so called?): but, once their nature has been recognized, we no longer feel tempted to call them ‘true’ or to dispute about their truth or falsity.

In the cases so far considered the model ‘This is red’ breaks down because the ‘statements’ assimilated to it are not of a nature to correspond to facts at all—the words are not descriptive words, and so on. But there is also another type of case where the words *are* descriptive

words and the 'proposition' does in a way have to correspond to facts, but precisely not in the way that 'This is red' and similar statements setting up to be true have to do.

In the human predicament, for use in which our language is designed, we may wish to speak about states of affairs which have not been observed or are not currently under observation (the future, for example). And although we *can* state anything 'as a fact' (which statement will then be true or false<sup>26</sup>) we need not do so: we need only say 'The cat *may be* on the mat.' This utterance is quite different from *tstS*—it is not a statement at all (it is not true or false; it is compatible with 'The cat may *not* be on the mat'). In the same way, the situation in which we discuss whether and state that *tstS* is *true* is different from the situation in which we discuss whether it is *probable* that *S*. *Tst* it is probable that *S* is out of place, inept, in the situation where we can make *tstST*, and, I think, conversely. It is not our business here to discuss probability: but is worth observing that the phrases 'It is true that' and 'It is probable that' are in the same line of business,<sup>27</sup> and in so far incompatibles.

## 7

In a recent article in *Analysis*, Mr. Strawson has propounded a view of truth which it will be clear I do not accept. He rejects the 'semantic' account of truth on the perfectly correct ground that the phrase 'is true' is not used in talking about *sentences*, supporting this with an ingenious hypothesis as to how meaning may have come to be confused with truth: but this will not suffice to show what he wants—that 'is true' is not used in talking about (or that 'truth is not a property of') *anything*. For it *is* used in talking about *statements* (which in his article he does not distinguish clearly from sentences). Further, he supports the 'logical superfluity' view to this extent, that he agrees that to say that *ST* is not to make any further assertion at all, beyond the assertion that *S*: but he disagrees with it in so far as he thinks that to say that *ST* *is* to *do* something more than just to assert that *S*—it is namely to *confirm* or to *grant* (or something of that kind) the assertion, made or taken as made already, that *S*. It will be clear that and why I do not accept the first part of this: but what of the

second part? I agree that to say that ST 'is' very often, and according to the all-important linguistic occasion, to confirm tstS or to grant it or what not; but this cannot show that to say that ST is not also and at the same time to make an assertion about tstS. To say that I believe you 'is' on occasion to accept your statement; but it is also to make an assertion, which is not made by the strictly performatory utterance 'I accept your statement.' It is common for quite ordinary statements to have a performatory 'aspect': to say that you are a cuckold may be to insult you, but it is also and at the same time to make a statement which is true or false. Mr. Strawson, moreover, seems to confine himself to the case where I *say* 'Your statement is true' or something similar—but what of the case where you state that S and I *say* nothing but '*look and see*' that your statement is true? I do not see how this critical case, to which nothing analogous occurs with strictly performatory utterances, could be made to respond to Mr. Strawson's treatment.

One final point: if it is admitted (*if*) that the rather boring yet satisfactory relation between words and world which has here been discussed does genuinely occur, why should the phrase 'is true' not be our way of describing it? And if it is not, what else is?

## Notes

1. It is sufficiently obvious that 'truth' is a substantive, 'true' an adjective and 'of' in 'true of' a preposition.
2. A likeness is true *to* life, but not true *of* it. A *word* picture can be true, just because it is *not* a picture.
3. Predicates applicable also to 'arguments,' which we likewise do not say are true, but, for example, valid.
4. Peirce made a beginning by pointing out that there are two (or three) different senses of the word 'word,' and adumbrated a technique ('counting' words) for deciding what is a 'different sense.' But his two senses are not well defined, and there are many more—the 'vocal' sense, the philologist's sense in which 'grammar' is the same word as 'glamour,' the textual critic's sense in which the 'the' in l. 254 has been written twice, and so on. With all his 66 divisions of signs, Peirce does not, I believe, distinguish between a sentence and a statement.
5. 'Historic' does not, of course, mean that we cannot speak of future or possible statements. A 'certain' speaker need not be any definite speaker. 'Utterance' need not be public utterance—the audience may be the speaker himself.



6. 'The same' does not always mean the same. In fact it has no meaning in the way that an 'ordinary' word like 'red' or 'horse' has a meaning: it is a (the typical) device for establishing and distinguishing the meanings of ordinary words. Like 'real,' it is part of our apparatus *in* words for fixing and adjusting the semantics of words.

7. Inverted commas show that the words, though uttered (in writing), are not to be taken as a statement by the utterer. This covers two possible cases, (i) where what is to be discussed is the sentence, (ii) where what is to be discussed is a statement made elsewhere in the words 'quoted.' Only in case (i) is it correct to say simply that the token is doing duty for the type (and even here it is quite incorrect to say that 'The cat is on the mat' is the *name* of an English sentence—though possibly *The Cat is on the Mat* might be the title of a novel, or a bull might be known as *Catta est in matta*). Only in case (ii) is there something true or false, viz. (not the quotation but) the statement made in the words quoted.

8. Both sets of conventions may be included together under 'semantics.' But they differ greatly.

9. 'Is of a type with which' means 'is sufficiently like those standard states of affairs with which.' Thus, for a statement to be true one state of affairs must be *like* certain others, which is a natural relation, but also *sufficiently* like to merit the same 'description,' which is no longer a purely natural relation. To say 'This is red' is not the same as to say 'This is like those,' nor even as to say 'This is like those which were called red.' That things are *similar*, or even 'exactly' similar, I may literally see, but that they are the *same* I cannot literally see—in calling them the same color a convention is involved additional to the conventional choice of the name to be given to the color which they are said to be.

10. The trouble is that sentences contain words or verbal devices to serve both descriptive and demonstrative purposes (not to mention other purposes), often both at once. In philosophy we mistake the descriptive for the demonstrative (theory of universals) or the demonstrative for the descriptive (theory of monads). A sentence as normally distinguished from a mere word or phrase is characterized by its containing a minimum of verbal demonstrative devices (Aristotle's 'reference to time'); but many demonstrative conventions are nonverbal (pointing, etc.), and using these we can make a statement in a single word which is not a 'sentence.' Thus, 'languages' like that of (traffic, etc.) *signs* use quite distinct media for their descriptive and demonstrative elements (the sign on the post, the site of the post). And however many verbal demonstrative devices we use as auxiliaries, there must *always* be a nonverbal *origin* for these coordinates, which is the point of utterance of the statement.

11. I use the following *abbreviations*:

S for the cat is on the mat.

ST for it is true that the cat is on the mat.

tst for the statement that.

I take tstS as my example throughout and not, say, tst Julius Caesar was bald or tst all mules are sterile, because these latter are apt in their different ways

to make us overlook the distinction between sentence and statement: we have, apparently, in the one case a sentence capable of being used to refer to only one historic situation, in the other a statement without reference to at least (or to any particular) one.

If space permitted other types of statement (existential, general, hypothetical, etc.) should be dealt with: these raise problems rather of meaning than of truth, though I feel uneasiness about hypotheticals.

12. If this is what was meant by “‘It is raining’ is true if and only if it is raining,” so far so good.

13. It takes two to make a truth. Hence (obviously) there can be no criterion of truth in the sense of some feature detectable in the statement itself which will reveal whether it is true or false. Hence, too, a statement cannot without absurdity refer to itself.

14. ‘It is true that S’ and ‘It is a fact that S’ are applicable in the same circumstances; the cap fits when there is a head it fits. Other words can fill the same role as ‘fact’: we say, e.g., ‘The situation is that S.’

15. We could use ‘nuts’ even now as a codeword: but a code, as a transformation of a language, is distinguished from a language, and a codeword dispatched is not (called) ‘true.’

16. Berkeley confuses these two. There will not be books in the running brooks until the dawn of hydrosentiments.

17. A statement may refer to ‘itself’ in the sense, for example, of the sentence used or the utterance uttered in making it (‘statement’ is not exempt from all ambiguity). But paradox does result if a statement purports to refer to itself in a more full-blooded sense, purports, that is, to state that it itself is true, or to state what it itself refers to (‘This statement is about Cato’).

18. And ‘by implication’ tstST asserts something about the making of a statement which tstS certainly does not assert.

19. This is not quite fair: there are many legal and personal reasons for holding two trials—which, however, do not affect the point that the issue being tried is not the same.

20. Not *quite* correct, because tstST is only in place at all when tstS is envisaged as made and has been verified.

21. *Unum, verum, bonum*—the old favorites deserve their celebrity. There is something odd about each of them. Theoretical theology is a form of onomatolatriy.

22. The following two sets of logical axioms are, as Aristotle (though not his successors) makes them, quite distinct:

- (a) No statement can be both true and false.  
No statement can be neither true nor false.
- (b) Of two contradictory statements—  
Both cannot be true.  
Both cannot be false.

The second set demands a definition of contradictories, and is usually joined with an unconscious postulate that for every statement there is one and only one other statement such that the pair are contradictories. It is doubtful how far any language does or must contain contradictories, however defined, such as to satisfy both this postulate and the set of axioms (b).

Those of the so-called 'logical paradoxes' (hardly a genuine class) which concern 'true' and 'false' are *not* to be reduced to cases of self-contradiction, any more than 'S but I do not believe it' is. A statement to the effect that it is itself true is every bit as absurd as one to the effect that it is itself false. There are *other* types of sentences which offend against the fundamental conditions of all communication in ways *distinct from* the way in which 'This is red and is not red' offends—e.g., 'This does (I do) not exist,' or equally absurd 'This exists (I exist).' There are more deadly sins than one; nor does the way to salvation lie through any hierarchy.

23. To be false is (not, of course, to correspond to a nonfact, but) to miscorrespond with a fact. Some have not seen how, then, since the statement which is false does not describe the fact with which it miscorresponds (but misdescribes it), we know which fact to compare it with: this was because they thought of all linguistic conventions as descriptive—but it is the demonstrative conventions which fix which situation it is to which the statement refers. No statement can state what it itself refers to.

24. Here there is much sense in 'coherence' (and pragmatist) theories of truth, despite their failure to appreciate the trite but central point that truth is a matter of the relation between words and world, and despite their wrongheaded *Gleichschaltung* of all varieties of statemental failure under the lone head of 'partly true' (thereafter wrongly equated with 'part of the truth'). 'Correspondence' theorists too often talk as one would who held that every map is either accurate or inaccurate; that accuracy is a singly and the sole virtue of a map; that every country can have but one accurate map; that a map on a larger scale or showing different features must be a map of a different country; and so on.

25. Austin pursues this line of thought further in *How to Do Things with Words* (Oxford: The Clarendon Press, 1962), pp. 139ff.—Ed.

26. Though it is not yet in place to call it either. For the same reason, one cannot lie or tell the truth about the future.

27. Compare the odd behaviors of 'was' and 'will be' when attached to 'true' and to 'probable.'

---

## A Realist Conception of Truth

William P. Alston

### 1 Introduction

In this essay I will set out what I call a “realist conception of truth” and defend it, insofar as that is required. The basic idea is a simple and familiar one. A statement, for example, is true if and only if (iff) what the statement is about is as the statement says it to be (more soberly, as the one making the statement says it to be). The statement that this room is lit is true iff what the statement is about, this room, is as it is said to be in making that statement, namely, lit. More succinctly, *the statement that this room is lit is true iff this room is lit*. The “content” of the statement, what it states to be the case, gives us everything we need to specify what it is for the statement to be true. That, in essence, is the conception of (propositional) truth I wish to defend.<sup>1</sup> It has many distinguished antecedents, reaching back at least as far as Aristotle, who said in a famous passage of the *Metaphysics*, “To say of what is that it is and of what is not that it is not is true” (IV, 6, 1001b, 28). But though the basic idea is very simple, it is not so easy to know how best to formulate it.

I will say something about that task in a moment, though not as much as in my recent book (Alston 1996), but first I want to explain why I call this a *realist* conception. Though ‘realism’ is more commonly used for one or another metaphysical position, I find it appropriate to call this conception of truth ‘realist’. The reason is this. What it takes to render a statement true is something that is objective vis-à-vis that statement, namely, a fact involving what the statement is about. The truth value of the statement depends on how it is with “the world” “beyond” the statement rather than on some feature of the statement itself. In

particular, and looking forward to the main competitor of the realist conception, truth value does not depend on the epistemic status of the statement, whether it is justified, warranted, counts as an expression of knowledge, or coheres with some system or other. I will use the term *alethic realism* for the view that the realist conception of truth is the one that is ordinarily used in application to statements, beliefs, and propositions.

Now for more elaborated formulations. There are two things to determine. (1) What to take as truth-value (“T-value”) bearers; what sort of thing is true or false. (2) How do we say, in general, what, on this conception, it is for a T-value bearer to be true? I will take them in that order.

## 2 The Choice of Truth Bearers

My brief introductory remarks were in terms of *statements*. But since by far the most popular choice for T-value bearers in recent English-speaking philosophy is *sentences*, I must say a word as to why I do not go along with this. First, we must distinguish between sentence types and tokens. A sentence type is the sort of entity that can be uttered and heard on many different occasions. You and I both utter *the sentence* ‘I’m hungry’. I utter it, the same sentence, many times. Here we are speaking of a sentence type. Each of the utterances of the sentence is a different sentence token. Most discussions of the truth of sentences deal with types. When someone brings out the old chestnut, “‘Snow is white’ is true iff snow is white’, she is not speaking of some particular utterance of that sentence but rather of what is common to all those utterances. But there are decisive reasons against attributing T-values to sentence types. The most serious one concerns the radical underdetermination of reference by meaning in natural languages. You and I both say ‘The indicator on the dial is at 7’, but it may be that what I say is true and what you say is false. The dial I’m looking at reads ‘7’, but you have misread the dial you are looking at. What are we to say of the sentence type ‘The indicator on the dial is at 7’? If we regard it as a bearer of T-values, we will have to say that it is sometimes true and sometimes false. And it’s worse than that. Since many people utter this sentence at various times, the sentence

type is constantly changing its T-value. But there are strong reasons against thinking of T-value bearers as so unstable. If I want to know whether it is true that the dial I was looking at reads 7, it is so that I can use that reading in the testing of some hypothesis. If the T-value of a T-value bearer with which I am concerned were constantly changing, or even occasionally changing, I could not either include it or reject it as a bit of evidence. And the same is true of more practical matters. I am concerned with whether it is true that you are hungry. If it is the sentence type 'I'm hungry' that is in question, *that* may well not retain the same T-value long enough for me to prepare food for you. These elementary points have been ignored by many philosophers in this century, primarily, I speculate, because of their preoccupation with artificial, formalized "languages," in which, since it is just stipulated what the referent of each singular term is, the kind of problem just mentioned does not arise.

Recently such considerations have led many philosophers to switch to sentence tokens as T-value bearers. A particular token can be assigned a stable T-value *provided* the speaker has satisfied the requirements, referential and otherwise, for making a definite statement. An alternative is to continue to ascribe T-values to sentence types, but relative to various contextual features that serve to pin down singular reference and other respects in which one token can differ in T-value from other tokens of the same type. So they think of the type 'I'm hungry' as having one T-value relative to one speaker and time and another T-value relative to a different speaker and time. On both of these alternatives, things become much less clear cut than they were when sentence types were straightforwardly taken to be true or false. In both cases matters other than purely linguistic ones are brought into the T-value bearer: the speaker, time of utterance, contextual factors that determine the reference of an expression like 'the chair', and so on.

However, the sentential choice faces an much more fundamental problem, one that points the way to a superior alternative. So far as I can see, there is no ordinary, nontechnical practice of applying 'true' and 'false' either to sentence types or to sentence tokens. Ask someone innocent of Anglo-American philosophy whether the sentence *type* 'The chair is broken' (not what someone is asserting by a particular utterance of that sentence) is true or false, and see if he can understand the question.

As for tokens, an utterance like 'Is that sequence of sounds you just made with your vocal organs true or false?' has no natural interpretation available to each fluent speaker of the language by virtue of her linguistic competence. But there is a way of introducing such a practice, at least for sentence tokens, as well as for sentence types that don't vary in ways I have been illustrating. For the token, we consider what statement was made in issuing that token and take the token to enjoy the T-value of that statement. For stable types, like 'Snow is white', we consider what statement would normally be made by a standard use of that sentence. And that, in effect, is what people do who assign truth values to sentences. This was implicit in the account I gave above of how sentence types can be assigned T-values *relative to certain other factors*. We pick factors that will affect what *statement* would be made by a particular utterance of the sentence.

But note where this has brought us. To understand what it is for a sentence to be true, we have to use the notion of the truth of a statement. A sentence token is true iff the statement made by uttering that sentence is true. And so even if attributing T-values to sentences is a viable project, it is conceptually dependent on thinking of statements as T-bearers. Hence statements are more fundamental bearers of T-values.

But in what sense of 'statement'? 'Statement' is ambiguous between the *act* of stating and *what is stated*, the *content* of the statement. Similarly 'belief', the term for another prominent T-value bearer, is ambiguous between the *psychological state* of believing something and *what is believed*, the *content* of the belief. It seems clear that in both cases it is in the content sense that 'true' or 'false' applies. When I wonder whether Smith's statement that Clinton will address the nation this evening is true, my interest is in whether it is true *that Clinton will address the nation this evening*, rather than in some feature peculiar to Smith's speech act. And the same holds of beliefs. You say that you believe that Clinton is innocent of the charges brought by Paula Jones, and I say "Do you really think that's true?" What is the referent of 'that' here? Not your psychological state of belief, but *that Clinton is innocent of the charges*.

Now for the final step. The content of a belief or statement can be termed a *proposition*. The 'that' clauses we use to specify those contents

can also be used to individuate propositions. Just as we can speak of the statement that gold is malleable and the belief that gold is malleable, so we can refer to the proposition that gold is malleable by, so to say, detaching the proposition from its status as the content of a statement or belief (or hope, fear, doubt, wondering, or whatever) and hold it up for examination on its own. And just as statements or beliefs can be termed true or false, so can propositions—naturally, since when we engage in the former talk, it is *what is stated* or *what is believed* to which we attribute a T-value; that is, we attribute the T-value to the *proposition* that is the content of the act of stating or believing. Hence we can take *propositions* as the most fundamental bearers of truth values. Statements and beliefs have that status by virtue of the propositions that are their contents, and sentences have that status, if at all, by derivation from statements.

Talk of propositions often raises philosophical hackles, and if I were seriously to address issues concerning the ontological status of propositions, I would never get to my main concerns here. Propositions are variously construed as abstract entities with an independent (Platonic) mode of timeless existence, as sets of possible worlds, as states of affairs that might or might not obtain, as complexes with structures that mirror those of sentences, and so on. Insofar as I have a view on such matters, it is Aristotelian rather than Platonist in that I think that the basic ontological locus of propositions is the acts of stating and the “propositional” attitudes in which they figure as contents. But for present purposes I sidestep all such questions. I take it that if one knows how to use ‘that’ clauses to specify statements, propositional psychological attitudes, and propositions, one has all the working grasp of the notion of propositions one needs to talk and think intelligibly of propositions as the basic bearers of T-values. The ontological chips may be left to fall where they may.

### 3 How to Formulate the Realist Conception of Truth

The next issue concerns how to give a general formulation of the realist conception of what it is for a proposition to be true. The initial rough formulation of the truth of a statement as dependent on *what the statement is about* seems less felicitous when applied to propositions as such. But the gut insight is the same. The proposition that this room is lit is



true iff the room is lit. Underlying this, and any other formulation concerning a particular proposition, is a general schema, which I will call the T-schema.

(1) The proposition that  $p$  is true iff  $p$ .

The similarity to Tarski's famous "equivalence of the form (T)" will not have escaped your notice.

(2)  $X$  is true iff  $p$ .

But the differences are equally significant.

- Schema (2) is about sentences, while schema (1) is about propositions.
- Schema (2), unlike schema (1), is designed for use with artificial, formalized "languages."
- Schema (2) is to be read in terms of material equivalence. Schema (1) is to be so understood that any substitution instance of the schema is a necessary, conceptual, analytic truth.<sup>2</sup>

Now (1), being a schema rather than a definite proposition, does not amount to a thesis about what it is for a proposition to be true. Nevertheless, it contains the seed of such a thesis. The simplest way to develop the seed into a full blown plant is to use substitutional quantification to give it a universal generalization.

(3) It is a necessary, conceptual, analytic truth that ( $p$ ) the proposition that  $p$  is true iff  $p$ .

This is not objectual quantification, but not because propositions are not "objects." Even if they are, it still doesn't count as objectual quantification, because the variables cannot be replaced by singular referring expressions that pick out objects. They must be replaced by declarative sentences. (Some philosophers try to treat declarative sentences as referring expressions, but so much the worse for them.) And substitutional quantification makes many philosophers nervous. Since I am not among them, I have no objection to treating (3) as a general statement of the position. But for those who do find it objectionable, there are alternatives. We can convey the same conception of propositional truth by going metalinguistic as in schema (4):

(4) Any substitution instance of (1) is a necessary, conceptual, analytic truth.

Anyone who realizes the necessary, conceptual, analytic truth of any substitution instance of the T-schema has grasped the realist conception of truth.

Note that none of this can claim to be a definition of truth, in the sense of a synonym of 'true' that can be substituted for 'true' whenever it occurs as a predicate of propositions. Even the most explicit formulation, (3), is not even a contextual definition. For even though, as I take it, it is true just by virtue of the meaning of 'true' that for any  $p$ ,  $p$ 's being the case is a necessary and sufficient condition for the truth of  $p$ , I am not willing to admit, as "deflationists" claim, that 'It is true that gold is malleable' is synonymous with 'Gold is malleable'. For the former contains a concept lacking in the latter, namely, the concept of truth. And it seems clear to me that one could understand 'Gold is malleable' (or at least simpler proposition-expressing sentences like 'The dog is scratching itself') without possessing the concept of truth. What I claim for (3) and (4) is that they are effective ways of bringing out the concept of propositional truth, in that any one who accepts them is thereby in possession of that concept.

#### **4 A Defense of the Realist Conception of Truth**

When I try to reflect on the question of why I accept this way of bringing out what it is for a proposition to be true, I find it difficult to know how to answer. For (3) and (4) seem to me to be miserable truisms, which no one who fully realized what he was saying would deny. If someone should say, "There is no doubt that oil is thicker than water, but it is not at all clear to me that it is true that oil is thicker than water," how should we respond? I would respond by saying that if the speaker had his mind on what he was saying and had no difficulty with the other terms of the utterance like 'doubt' and 'clear', then we must judge that he is deficient in his grasp of 'true', that he simply does not have the ordinary concept of propositional truth. It simply doesn't make sense to say 'Oil is thicker than water, but it is not true that oil is thicker than water'.

But if this is so truistic, why should I have devoted a sizable book to laying it out and defending it? As for laying it out, the difficulty of identifying the best way of formulating it may be a sufficient reason. But as

for defending it, the reason is that many contemporary thinkers deny it, or at least take positions that appear to be inconsistent with it. As Cicero once wrote, there is nothing so absurd that it may not be found in the books of the philosophers. And if it weren't for the fact that philosophers, especially the cleverest among them, are given to espousing and defending what seem to be obviously false positions, the more sensible among us might be at a loss as to how to spend our time. (Or perhaps we would find more useful pursuits!) Against that background I feel it is not a waste of time to exhibit clearly and defend the ordinary way of understanding propositional truth.

A word is in order concerning the relation of the realist conception of truth to various forms of metaphysical realism and antirealism. Since metaphysical realism is a large and sprawling territory, I cannot properly enter into it in this essay. But I will just stick my toe in. My account of truth is neutral between the historically prominent metaphysical debates between realists and their opponents. The basic point can be put this way. The metaphysical realist and the metaphysical antirealist differ as to *what* propositions are true or false, but they need not differ as to *what it is* for a proposition to be true or false. This is obvious with respect to what we might call "parochial" realisms and antirealisms, realism with respect to a certain putative domain of reality: properties, propositions, and other "abstract objects"; physical objects; theoretical entities; moral properties; aesthetic values; God. The "realist" and the phenomenalist about physical objects can agree that their claims are true or false, depending on whether what they are talking about is as they say it to be. The same point can be made concerning realists and instrumentalists about theoretical entities, and concerning theists and atheists. I would even say that the more "objective" forms of idealism, typified by Berkeley, can accept a realist account of truth. On a Berkeleyan idealism, a certain physical fact, e.g., that there is a spruce tree in my front yard, turns out to be a fact about the mind of God. But whether it is true that there is a spruce tree in my front yard depends on whether the mind of God is organized as that proposition (interpreted in a Berkeleyan way) would have it organized.

I am prepared to go further and hold that even an ontological relativism like that espoused by Putnam or Goodman is compatible with

realism about truth. On these views, what seems to be a purely objective fact (say about my spruce tree) obtains only relative to a certain conceptual scheme, one to which there are viable alternatives. But if we put the relativity into the content of propositions rather than in the concept of truth (something that Putnam sometimes denies), then we can say, in the spirit of the T-schema, that it is true that, relative to scheme C, there is a spruce tree in my front yard iff relative to scheme C there is a spruce tree in my front yard.<sup>3</sup>

To be sure, what a given philosopher calls ‘metaphysical realism’ may include a commitment to a realist conception of truth, and when that philosopher opposes “metaphysical realism,” that opposition can be expected to include a rejection of the realist conception of truth. We find this exemplified in, for example, Putnam 1981. But from my point of view, in that book Putnam has linked together what I would regard as a distinctively metaphysical realism (“The world consists of some fixed totality of mind-independent objects”) and a realist view of truth (“Truth involves some sort of correspondence relation between words or thought-signs and external things and sets of things”) (1981, 49).<sup>4</sup> If we separate out the strictly metaphysical claim, we find that one can deny that without denying the realist conception of truth.

## 5 Does the Account Go Far Enough?

Even if one were prepared to admit that my account is accurate, as far as it goes, one might maintain that it does not go far enough. To bring out the basis for this reaction, I need to identify how my account is *minimal*. Its minimality consists in its undertaking the fewest commitments compatible with identifying the concept of truth. It is restricted to bringing out how the truth of a proposition depends, to vary the expression a bit, on whether its content is “actualized” or “realized” in the “world.” In the terms I used above, it is confined to affirming the conceptual truth of all instantiations of the T-schema, and anything equivalent to that. But traditionally accounts of truth that are of a generally realist cast have been more ambitious. They have taken truth to consist in some kind of *correspondence* between a proposition and a fact. This is typically spelled out by specifying a kind of structural isomorphism that must hold between

a proposition and a fact if the proposition is to be true. And my account is silent about all that. I take this to be a virtue, but the present objection deems it a vice. So which is it?

My answer is that it is both, and to sustain that answer I need to make a distinction. The distinction I need is one that, in other cases, has been prominent in philosophy lately: that between the *concept* of *P* and the *property* of *P*. To take an familiar example, the ordinary concept of water is something like *a stuff that is liquid (in certain temperature ranges), tasteless and colorless when pure, what falls in rain, what is in oceans, lakes, and streams, etc.* The ordinary, pretheoretic concept contains no specification of chemical composition, much less finer physical structure. But empirical investigation has revealed that the *property* of being (pure) water, the property of belonging to that *natural kind*, is *having the chemical constitution*  $H_2O$ . This is a feature of the kind that, while compatible with the features represented in the concept, goes beyond them to a significant extent. The same distinction can be made here. Even if the ordinary concept of truth is adequately picked out by my T-schema-based account, it may be that further investigation will reveal additional features of *what truth is*, what the property is whose possession makes a proposition true. To be sure, this case differs from the water case in that the investigation will not be empirical in the same way. It will consist, rather, of reflection on a proposition's being true and of an attempt to specify what is necessarily involved in that. I have just hinted at a direction such an investigation might take: exploring the structures of propositions and facts and spelling out what it takes for the right kind of "match." In these terms I deny that my account does not "go far enough" to identify the *concept*, but I agree that it does not go so far as to spell out features of the *property* that go beyond that. My position on the concept is not committed to the success of any such further characterization, but it is not committed to its failure either.

Moreover, my realist conception of truth is by no means neutral between different ideas as to what further features the property might have. As I will argue below, it sorts ill with accounts of the property of truth in terms of epistemic features of T-value bearers. On a more positive note, it seems clear that the T-schema suggests a correspondence

theory of the property; indeed, such a theory seems to be implicit in the schema. The T-schema naturally gives rise to the idea that a proposition is made true by a fact. The dictum that *the proposition that lemons are sour is true iff lemons are sour* is naturally read as saying that the proposition is made true by the *fact* that lemons are sour (rather than that lemons are made sour by the truth of the proposition). We could embody that idea in what might be called a minimalist form of a correspondence theory.

(5) (*p*) the proposition that *p* is true iff it is a fact that *p*.

This is, at most, an inchoate form of correspondence theory because it does nothing to spell out how a fact has to be related to a proposition to make it true. But the T-schema does *exhibit* what it is about a particular fact that makes a particular proposition true; it does so by using the same ‘that’ clause to specify both. We might put this by saying that it is an identity of “content” that makes that fact, rather than some other, the truth maker for that proposition. And this talk of identity of content has brought us to the verge of a full-blown correspondence theory, which would go into what the “content” of propositions and of facts consists in—an attempt that would presumably lead to a specification of something like a structural isomorphism between a proposition and the truth-making fact.

## 6 Epistemological Objections to the Realist Conception

Now I want to consider a widespread kind of objection to a realist conception of truth, what I call an “epistemological” objection. By this I mean not an argument that an epistemic account of the concept is superior to a realist one, but rather an argument that is concerned with the epistemology of truth, with what it takes to tell whether a given proposition is true. The argument is that this is impossible on a realist conception of truth.

The argument exists in many versions. A prominent one depends on the assumption that determining that a proposition is true, on the realist conception, requires “comparing a proposition with a fact,” and it is argued (or more frequently, just assumed) that this is impossible.<sup>5</sup>

Here are some dicta to that effect (in terms of statements or beliefs, rather than propositions, but we have seen to how to translate back and forth):

Each statement may be combined or compared with other statements, e.g., in order to draw conclusions from the combined statements, or to see if they are compatible with each other or not. But statements are never compared with a “reality,” with “facts.” None of those who support a cleavage between statements and reality is able to give a precise account of how a comparison between statements and facts may possibly be accomplished, and how we may possibly ascertain the structure of facts. (Hempel 1935, 50–51)

If meanings are given by objective truth conditions there is a question how we can know that the conditions are satisfied, for this would appear to require a confrontation between what we believe and reality; and the idea of such a confrontation is absurd. (Davidson 1986, 307)

Justification is a matter of accommodating beliefs that are being questioned to a body of accepted beliefs. Justification always terminates with other *beliefs* and not with our confronting raw chunks of reality, for that idea is incoherent. (Williams 1977, 112)

Neither the claim that to tell that a proposition is true, on a realist account, requires comparing a proposition and a fact, nor the claim that this is impossible, is supported by any argument here. And why should we suppose that I have to make any such comparison to discover that a proposition is realistically true? In reflecting on this question, we discover an important distinction between ways of understanding the requirement. It is susceptible of an innocuous interpretation in which whenever I recognize that it's true that this room is lit just by recognizing that this room is lit, I have carried out a comparison of proposition and fact. If *that's* all it amounts to, it is unsurprising that philosophers like those just quoted mount no argument for its impossibility. But presumably they have in mind something more ambitious and (allegedly) more difficult. Thus, it is sometimes made explicit that it is a maximally *direct* awareness of facts and their relation to propositions, along with the epistemic statuses of infallibility and indubitability customarily associated with direct awareness, that is said to be both required by the realist conception and to be impossible. This is suggested by Davidson's and Williams's use of the metaphor of *confrontation*. It is more explicit in Rorty's 1979 attack on a realist understanding of truth as presupposing an indefensible

account of the mind as a “mirror of nature,” as capable of unmediated, foolproof awarenesses of extramental fact.

But why suppose that even an explicit, conscious comparison of fact and proposition requires that one be *immediately* aware of the fact in question? Why wouldn’t it be enough to have any sort of knowledge of fact, whether immediate, inductively derived, based on inference to the best explanation, or whatever? So long as we know that the fact obtains, why should it matter how we get that knowledge?

In any event, if we think of a kind of awareness that is properly termed ‘direct’, even if not infallible and indubitable, such as perception of the immediate environment as viewed by direct realists about perception, why isn’t that possible and why can’t that be a basis for determining whether facts and propositions match? Here is a spirited defense of that possibility by Moritz Schlick, in response to the 1935 article by Hempel, from which I quoted above.

I have been accused of maintaining that statements can be compared with facts. I plead guilty. I have maintained this. But I protest against my punishment. I have often compared propositions to facts; so I had no reason to say that it couldn’t be done. I found, for instance, in my Baedeker the statement: “This cathedral has two spires,” I was able to compare it with “reality” by looking at the cathedral, and this comparison convinced me that Baedeker’s assertion was true. . . .

Perhaps you say: “But if we analyze the process of verification of Baedeker’s assertion we shall find that it amounts to a comparison of propositions.” I answer: whatever the result of your analysis may be, at any rate we can distinguish between cases in which a written, printed or spoken sentence is compared with some other written, printed or spoken sentence, and cases like our example, where a sentence is compared with the thing of which it speaks. And it is this latter case which I took the liberty of describing as a “comparison of a proposition with a fact.” . . .

You insist that a statement cannot or must not be compared to anything but statements. But why? It is my humble opinion that we can compare anything to anything if we choose. Do you believe that propositions and facts are too far removed from each other? Too different? Is it a mysterious property of propositions that they cannot be compared with anything else? That would seem to be a rather mystical view. (Schlick, in Macdonald 1954, 232–235)

Against this eminently commonsensical protest by Schlick there is a serious argument from the nature of perceptual cognition, an argument that was prominent in absolute idealism and that has enjoyed a recent



revival of influence. This argument picks up on Schlick's surmise that his opponent will claim that his perceptual verification amounts to a comparison between propositions. We find it, to pick one source out of a crowd, in Blanshard 1939. There, in responding to a claim like Schlick's, he writes as follows:

It [the position Blanshard opposes] assumes that, corresponding to our judgment, there is some solid chunk of fact, directly presented to sense and beyond all question, to which thought must adjust itself. And this "solid fact" is a fiction. What the theory takes as fact and actually uses as such is another judgement or set of judgements, and what provides the verification is the coherence between the initial judgement and these. (Blanshard 1939, vol. 2, p. 228)

This is a form of the currently popular view that perceptual awareness of objects is conceptually structured and, in stronger forms like that of Blanshard's, propositionally structured in such a way as to involve judgments with the propositional content in question. Hence Blansard takes it that one can't see a cardinal without judging it to be such and such (not necessarily to be a cardinal). And from that he draws the conclusion that the supposed external fact of the cardinal sitting on a branch is really itself a judgment (statement, proposition), rather than a fact correspondence with which could render the judgement true.

But, of course, even if all perceptual awareness is propositionally structured, and even if it all involves judgement (which I do not admit for a moment), it would *not* follow that there is nothing to the perceptual awareness of a cardinal but a judgement. Seeing a cardinal is obviously different from merely judging that a cardinal is there. And that difference reflects something in the perception that is in addition to judgement, some kind of distinctively perceptual awareness of what any judgement that may be involved is about. And so the thesis that perception has a pervasively propositional structure leaves open the possibility that this distinctively perceptual awareness of objects might constitute a presentation, even a direct presentation, of extrajudgmental fact. But though I don't think that the proponents of this objection to a realist conception of truth have closed off the possibility of a direct awareness of extramental facts, I don't want to rest my case on that highly controversial claim. Instead what I take to be my strongest point is the earlier one: that even if no such direct awareness is possible, one can "compare" propo-

sition and fact, provided one has knowledge of each, whether that knowledge is direct or not.

Finally, I want to look at a more general epistemological argument against a realist conception of truth. This one is based on the supposition that it is essential to realism to construe truth value as determined by a reality that is external to our knowledge, not just in the sense of being other than our knowledge, but also in the sense of being *inaccessible* to it. There is a forthright statement of this position in Horwich 1982. There he characterizes “metaphysical realism” as the doctrine that “the concept of truth involves a primitive non-epistemic idea—for example, ‘correspondence with reality’.... Truth is held to be a genuine property of certain propositions ... and ... the goal that motivates our standards of justification and our verification procedures” (1982, 182). This is along the same lines as my alethic realism. Horwich goes on to say,

The respect in which metaphysical realism is committed to autonomous facts is ... radical. It concerns the adequacy of the canons of justification implicit in scientific and ordinary linguistic practice—what reason is there to suppose that they guide us towards the truth? This question, given metaphysical realism, is substantial and, I think, impossible to answer; and it is this gulf between truth and our ways of attempting to recognize it which constitutes the respect in which facts are radically autonomous. Assuming a grasp of propositions, and knowledge of *what it is* for them to have the property of metaphysical truth, it is far from clear how we could derive the ability to recognize *when* this property applies. Indeed, it is our total inability to see how this problem might be solved which should lead us to reject metaphysical realism.... Thus metaphysical realism involves to an unacceptable, indeed fatal, degree the autonomy of facts: there is from that perspective no reason to suppose that scientific practice provides even the slightest clue to what is true. (1982, 185–186)

Needless to say, Horwich is free to define ‘metaphysical realism’ in any way he pleases. What he is not free to do is first to define the view in terms of a nonepistemic concept of truth that takes it to be something like correspondence with fact, and then to attribute to the view so defined a representation of facts whereby we have no way of determining what the facts are, without giving any reasons for supposing that this view of truth is committed to the latter.<sup>6</sup> And he gives no reason for supposing that “metaphysical realism,” as he defines it, is committed to taking the truth-making facts to be cognitively inaccessible. It is sheerly arbitrary to burden the view with such crippling consequences.

Another attempt to saddle realism about truth with the doctrine that facts are inaccessible is found in Rorty's essay, "The World Well Lost" in Rorty 1982, in which he presents this extreme conception as the only alternative to regarding "the world" as consisting of the beliefs we take to be firmly established and immune from doubt. Referring to the Davidsonian position that it is necessary that most of our beliefs are true, he writes,

If one accepts the Davidson-Stroud position, then "the world" will just be the stars, the people, the tables, and the grass—all those things which nobody except the occasional "scientific realist" philosopher thinks might not exist. So in one sense of 'world'—the sense in which (except for a few fringe cases like gods, neutrinos, and natural rights) we now know perfectly well what the world is like and could not possibly be wrong about it—there is no argument about the point that it is the world that determines truth. . . . But this is, of course, not enough for the realist. What he wants is precisely what the Davidson-Stroud argument prevents him from having—the notion of a world *so* "independent of our knowledge" that it might, for all we know, prove to contain none of the things we have always thought we were talking about. This notion of the world must be the notion of something *completely* unspecified and unspecifiable—the thing in itself, in fact. To sum up the point, I want to claim that "the world" is either the purely vacuous notion of the ineffable cause of sense and goal of intellect, or else a name for the objects that inquiry at the moment is leaving alone: those planks in the boat which are at the moment not being moved about. (1982, 14–15)

Here Rorty, the master of caricature, is exercising his art. If you are not content to construe "the world," "reality," as consisting of beliefs that are taken as firmly established (or as needing no establishment), then the only alternative is a world of which we can know nothing. But as soon as this disjunction is formulated, it can be seen to be obviously not exhaustive. Why couldn't the world be made up of facts that are what they are, independent of our cognitive successes and failures, without these facts being all of them inaccessible to our knowledge? Why foist onto the realist, who takes the truth of *p* to depend solely on whether it is the case that *p*, the commitment to *p*'s being unknowable? Isn't there room for the category of *might or might not be known*, as well as the categories of *known* and *unknowable*?

You will note that Rorty's argument attacks a form of metaphysical realism properly so called that takes the facts on which (many) propositions depend for their truth value to be what they are independent of our

cognitive dealings with them. Rorty, like many philosophers, makes no sharp distinction between such metaphysical realism and realism about truth. (Indeed, Rorty associates the T-schema with his *alternative* to realism about truth!) By contrast, I have already dissociated my realism about truth from any commitment to such a metaphysical position, though the two positions do have an affinity for each other. Nevertheless, Rorty's line of argument can be turned against a realist construal of truth, and that is why I introduced it here. This is because Rorty presents the view of reality as cognitively inaccessible by us as the only alternative to a coherence theory of truth, and this leaves no room for a realist account of truth on which the facts that make propositions true are often accessible. Hence it is not irrelevant to the defense of my view to point out the defects in arguments that purport to show that any alternative to a coherence account of truth makes truth undiscoverable.

## 7 Epistemic Conceptions of Truth

The main alternatives to a realist conception of truth are *epistemic* conceptions, which identify truth with some positive epistemic status of T-value bearers. This positive epistemic status is variously identified with membership in a maximally comprehensive and coherent system (as with the absolute idealist views of truth as coherence that were prominent in the late 1800s and early 1900s), with what, in Peirce's well-known formulation, "is fated to be ultimately agreed to by all who investigate," with what, in Dewey's phrase, is "instrumental to an active reorganization of the given environment, a removal of some specific trouble or perplexity," and more recently in Putnam 1981, with being such that it "would be justified in epistemically ideal conditions." To give focus to this brief discussion, I will concentrate on Putnam's view, abbreviating his candidate for truth as "ideal justifiability" (IJ) and his view as the "ideal justifiability conception" (IJC).

Though epistemic conceptions of truth have been attractive to many, I believe that they are among the few widely held philosophical positions that can be definitively refuted. I will give brief presentations of four arguments against them.<sup>7</sup>

## i

The first is an “extensional” argument to the effect that truth and IJ are not completely coextensive. That is the case if there are propositions that are true but not IJ, or IJ but not true. I will concentrate here on the former possibility. If I were to try to give particular examples, it could be objected that I can hardly be confident that a belief is true without supposing myself to be justified in accepting it and supposing that this justification would hold up no matter how improved my epistemic situation. Even if that is so, it would not follow that a belief could not *be* true without being IJ. But rather than continuing that argument, I prefer to proceed more indirectly. I will consider how plausible it is to hold that there are true propositions that would not be justifiable in an ideal epistemic situation.

The most extreme candidates would be propositions such that nothing that tells for or against their truth is cognitively accessible to human beings, even in principle. I need not restrict myself here to propositions we are able to envisage. May there not be states of affairs, or even entire realms or aspects of reality, that are totally inaccessible to human cognition? If so, propositions to the effect that such states of affairs obtain will be true, even though no beliefs or statements bearing those propositions as their content would be justifiable in an epistemically ideal situation.

But how plausible is it that there *are* realms or aspects of reality that are in principle inaccessible to human cognition? There are considerations that render it quite plausible. Think of the limitations of our cognitive powers—limitations on our storage and retrieval capacity, on the amount of data we can process simultaneously, on the considerations we can hold together in our minds at one moment, on the complexity of propositions we are capable of grasping. Isn’t it highly likely that there are facts that will forever lie beyond us just because of these limitations? And it is not just our finitude; there is also what we might call our “particularity.” The cognitive design of human beings represents only one out of a large multitude of possible designs for cognitive subjects, even for embodied cognitive subjects as finite as we are, leaving out of account angels and God. It seems clear that there could be corporeal cognitive subjects with forms of sensory receptivity different from ours, with sen-

sitivity to different forms of physical energy. There could be subjects with different innate cognitive tendencies, propensities, and hard-wired beliefs and concepts. There could be subjects who reason in patterns different from those we employ. All this strongly suggests that there are many facts accessible to cognizers with radically different hardware and software but totally inaccessible to us.

One possible response to this objection would be to make the IJC range over cognitive subjects generally. Truth would then be identified with justifiability for some cognitive subjects or other in situations that are the most ideal for those subjects. And if there are still true propositions unenvisageable by any actual subjects, we could make the conception range over possible subjects as well. These modifications would certainly take the sting out of the present objection. But it would also take much of the sting out of the IJC. If we survey the reasons that have been given for an epistemic definition of truth, we will see that they are heavily anthropocentric. Dummett's arguments for a verificationist conception of truth, for example, depend on considering what sorts of truth conditions *we* could learn to attach to sentences. James and Dewey are preoccupied with how *we* judge beliefs to be true or false and with the functions beliefs we call true play for *us* in *our* lives. And Putnam writes, "A true statement is a statement that a rational being would accept on sufficient experience of the kind that is actually possible for beings *with our nature* to have" (1981, 64). And in any event, we would have to restrict consideration to finite cognitive subjects. If an omniscient deity were brought into the picture, the position would lack the antirealist bite it is designed to have.<sup>8</sup> Realism should have no hesitation in recognizing that a necessary condition of the truth of a proposition is that it would be known (accepted, believed, etc.) by an omniscient cognitive subject. And with the restriction to finite subjects in place, we still have to take seriously the idea that some aspects of reality are inaccessible in principle to *any* subjects—actual or possible.

## ii

Another objection is that we can't spell out ideal justifiability without making use of the concept of truth. Hence the explication cannot go the other way without circularity. Here are two ways of seeing this.

First, what is meant by one's being (epistemically) justified in holding a certain belief? Most epistemologists who address this issue take a "truth-conducivity" position, according to which a belief is justified iff it is formed and/or held so as to make the belief likely to be *true*. There is a variety of suggestions as to what confers this likelihood: being based on adequate evidence, grounds, or reasons; being formed by the operation of a reliable belief-forming mechanism; cohering in the right sort of system; etc. But in taking these to be justification-conferring conditions, one supposes them to render the belief likely to be true. And the basic reason for this is that otherwise justification would not have the value for our cognitive endeavors that we take it to have. Laurence Bonjour puts the matter strongly:

Why should we, as cognitive beings, care whether our beliefs are epistemically justified? Why is such justification something to be sought and valued?... The following answer seems obviously correct.... What makes us cognitive beings at all is our capacity for belief, and the goal of our distinctively cognitive endeavors is *truth*: we want our beliefs to correctly and accurately depict the world.... The basic role of justification is that of a *means* to truth, a more directly attainable mediating link between our subjective starting point and our objective goal.... *If our standards of epistemic justification are appropriately chosen*, bringing it about that our beliefs are epistemically justified will also tend to bring it about... that they are true. If epistemic justification were not conducive to truth in this way, if finding epistemically justified beliefs did not substantially increase the likelihood of finding true ones, then epistemic justification would be irrelevant to our main cognitive goal and of dubious worth. (Bonjour 1985, 7–8)

But if that is the case, if epistemic justification is essentially truth-conducive, then unless 'justification' is being used in some different sense that would need explanation, being justified in an ideal epistemic situation would differ from being justified by ordinary, everyday standards only in that it is even more strongly indicative of truth than the latter. Hence we can't explain what is meant by an ideal epistemic situation without employing the concept of truth.

This point can be driven home by considering some alternatives to a truth-conducivity conception of justification. Foley (1987) holds that a belief is "epistemically rational" iff the believer would, on sufficient reflection, take there to be a conclusive argument for it (where in limiting cases the argument can be from itself to itself). This makes justification independent of truth by carrying out a considerable subjectivization of

the concept of justification. And this very feature makes justification, so construed, of doubtful cognitive value. If we allow an unrestrained variation in what a given individual would take, on considerable reflection, to be a conclusive argument for  $p$ , we may well wonder why anyone should be concerned to see to it that their beliefs are “epistemically rational.”

A somewhat more widespread approach is to tie justification to the lack of any violation of intellectual obligations or duties. Just I am justified in an action, e.g., resigning from my job with only two weeks notice, provided I violate no rule, regulation, or commitment in doing so, on this view, I am justified in believing that there is intelligent life elsewhere in the universe, provided my believing this is not the result of failures to conduct my intellectual operations as I should. To be sure, whether this way of thinking of justification is opposed to a truth-conducivity conception depends on what intellectual obligations we have. If they include an obligation to do what we can to believe what is true and to avoid believing what is false, this is not sharply opposed to a truth-conducivity view. But if we think of these obligations in some other way, e.g., as looking carefully for pro and con considerations on the issue, then this may or may not be closely connected with a likelihood of truth for the belief in question. And if it isn't, the question again arises as to why we should take justification in this sense to be an important value for our cognitive lives.

So truth already enters into the very conception of epistemic justification. Another way in which truth is presupposed in the understanding of IJC has to do with what makes an epistemic situation *ideal*. A natural understanding of this is in terms of the *ready availability of all relevant evidence* (reasons, considerations). Now to say that evidence is “available” is to say that one could come into possession of it. So the crucial notion is that of *possession* of evidence. The evidence itself will presumably consist of facts. What is it for the subject to *possess* those facts so as to make use of them in justifying a belief? The most obvious answer is that the subject comes to *know* them. But the notion of knowledge involves the notion of *truth*. (Knowledge is *true* belief that satisfies certain further conditions.) To avoid this conclusion, we would have to construe possession of the facts in terms of belief without mentioning



knowledge. But then we will have to require the beliefs to be *true*. Otherwise, one is not in possession of genuine evidence but only mistakenly supposes himself to be so. At this point the suggestion might be that possession of evidence consists in having *justified* beliefs. But we have already seen that the notion of justification involves the notion of truth. Moreover, this involves a different circle. The justification involved here obviously can't be justification by everyday standards. For by those standards, a belief can be justified in one situation and not in another, which means that there is no unique answer to what the relevant evidence is for a given target belief. Hence these will have to be beliefs that would be justified in *ideal epistemic circumstances*. But then we are in an even smaller circle. We define 'ideal justifiability' in terms of an ideal epistemic situation, but then we have to define such a situation in terms of what beliefs would be justified in an ideal epistemic situation!

I believe there are other points at which the concept of an ideal epistemic situation rests on the concept of truth, but sufficient unto the day is the difficulty thereof.

### iii

Next I would like to present an argument against the IJC that is based on the T-schema, the heart of alethic realism. According to the T-schema, or a suitable generalization thereof, the fact that sugar is sweet is conceptually both necessary and sufficient for its being true that sugar is sweet. And that would seem to leave no room for any *epistemic* necessary or sufficient conditions. Nothing more than *sugar's being sweet* is needed to make the proposition true, and nothing less would suffice. How, then, can some epistemic condition be conceptually necessary and/or sufficient?

My opponent can complain that I am begging the question by basing the argument on what my position takes to be conceptually necessary and sufficient for truth. But to see that the argument is free of special pleading, we only have to note that the T-schema is almost universally endorsed by epistemic theorists (though many of them, thinking in terms of sentences as T-value bearers, endorse the Tarskian version instead). Thus Putnam writes, "We could ... *keep* formal semantics (including 'Tarski-type' truth-definitions) ... and yet *shift* our notion of 'truth' over

to something approximating ‘warranted assertibility’” (Putnam 1978, 29). Thus epistemic theorists are anxious to square their position with the T-schema or its brethren. And that is not surprising if the T-schema is as obviously true as I have been claiming. And so in arguing from that schema against epistemic theorists’ construal of the concept of truth, I rely on what is common ground between us.

But, of course, there is a possible answer to my charge. An epistemic concept of truth like the IJC may not be incompatible with the T-schema. It wouldn’t be if IJC were conceptually equivalent with the realist necessary and sufficient condition for truth, namely, the corresponding fact. If *the belief that sugar is sweet being ideally justifiable* is itself conceptually equivalent to *sugar’s being sweet*, then, by the transitivity of conceptual equivalence, the former is likewise conceptually equivalent to its being true that sugar is sweet. I say this is a possible answer, but, so far as I can see, it is only abstractly conceivable. In particular, it is no epistemic possibility. What basis could there be for holding that the ideal justifiability of the belief that sugar is sweet is *conceptually* both necessary and sufficient for sugar’s being sweet. Even if they were extensionally equivalent, which I do not admit, why suppose that the concepts involved guarantee the equivalence. On the basis of linguistic intuition, it seems clear that we can consistently conceive of a fact (if not sugar’s being sweet, then some more recondite fact like the big bang’s being preceded by a collapse of a previous universe into a point) without its being ideally justifiable that the fact obtains. The most I can see to be even minimally plausible along this line is the following. On one type of absolute idealism, anything I can think of that is external to my current thought is some fully realized development of my thought, an “all comprehensive and fully realized whole” of which Anglo-American absolute idealism spoke, or, in more Hegelian terms, the culmination of the Absolute Spirit’s process of attaining full self-development. On such a view it would, if you like, be *metaphysically* impossible that sugar would be sweet without that judgment’s figuring in an all comprehensive and ideally coherent system of thought. But that would still not make that combination *conceptually* impossible. It would still not prevent us from consistently and intelligibly envisaging that sugar is sweet although it is not ideally justifiable that it is. And so

this suggested out for the epistemic theorist turns out to be a blind alley. And even if it weren't, I doubt very much that any contemporary advocate of an epistemic conception of truth would be willing to purchase acceptability for the view at the price of accepting absolute idealism.

iv

One final shot against IJC. We have seen that a common objection against a realist conception of truth is that it makes it impossible to determine truth values. And correspondingly, a main attraction of epistemic conceptions is that they avoid this disability of their rival. But ironically enough, the tables are turned against epistemic accounts on just this point. As soon as we make the epistemic conception strong enough to be at all plausible as an account of truth, it turns out that on that conception, it is *much more* difficult to determine truth values than it is on the realist conception. If we could identify truth with, say, justification by ordinary standards, then it would be easier to determine truth values than on the realist conception. But any such account would be palpably inadequate. Clearly, some beliefs justified by ordinary standards are false. For an epistemic account of truth to have any plausibility at all, we have to identify truth with some highly idealized epistemic status: membership in a *maximally* comprehensive and coherent system or what would be justified in *ideal* epistemic conditions. And having inflated the conception to that extent, it becomes extremely problematic whether a belief satisfies the condition. Who can say which of our present beliefs would still be justifiable if we had ready access to *all* relevant evidence? Whenever the topic is difficult or controversial, as with many issues in science, history, and philosophy, we are in no position to say with any assurance what position would be justified in the most ideal of circumstances. And where simpler matters are concerned, as with garden-variety perceptual judgments, we are in a much better position with a realist conception. Because of the severe problems of working out a determinate conception of ideal epistemic circumstances, it is much easier to determine that my computer is on now than it is to determine whether that belief would be justified in *ideal epistemic circumstances*. So the IJC doesn't deliver the goods for the sake of which it is sought. It makes truth values less accessible, not more.

## Notes

This essay is mostly a condensation of material to be found in chapters 1, 3, and 7 of Alston 1996, with a few additional twists. A closer ancestor is two keynote lectures delivered at the 1997 Wheaton Philosophy Conference on Truth and Realism.

1. There are other senses of 'true' in addition to the one I will be examining. We speak of a 'true friend', a 'true copy', 'being true to one's word', and so on. My interest is in the sense of 'true' that applies to beliefs, statements, propositions, and the like. When I want to emphasize this, I will use the term 'propositional truth', but even when I don't, this restriction should be understood.

2. Talk of *conceptual* or *analytic* truths is distasteful to those philosophers who deny that we can distinguish between what belongs to our *concept* of *P* and what is obviously true of *Ps*, or between what is true solely by virtue of the meanings of terms and what is true, at least in part, on other bases. I agree that it is often difficult, even impossible, to find a sharp line of division. But it also seems clear that there are many clear cases on both sides. The conceptual, analytic truth of instances of the T-schema is one of those clear cases.

3. See Alston 1996, chap. 6, and Lynch 1998 on this point.

4. This correspondence view of truth goes beyond my "minimalist" version of a realist conception of truth, but as I will suggest below, it is in the direction pointed out by my account.

5. This argument is usually directed against a correspondence theory of truth, but we can take it to be directed against my minimalist realist conception of truth as well. For, as just pointed out, my position holds that a proposition must share a content with a fact if it is to be true. And so it may be argued that my position, as much as a full-blown correspondence theory, implies that I can tell that a proposition is true only by telling that it corresponds with a fact.

6. Actually, Horwich says both this and that on metaphysical realism we have no reason for supposing that satisfying *ordinary criteria of justification* is likely to get us closer to the truth, which is not equivalent to its being impossible to know the truth makers.

7. The arguments are presented in more detail in Alston 1996, chap. 7.

8. See Plantinga 1982. Putnam 1981 identifies "metaphysical realism," which he opposes, with a "God's-eye point of view."

## References

- Alston, William P. 1996. *A Realist Conception of Truth*. Ithaca: Cornell University Press.
- Blanshard, Brand. 1939. *The Nature of Thought*. 2 vols. London: George Allen and Unwin.

- BonJour, Laurence. 1985. *The Structure of Empirical Knowledge*. Cambridge: Harvard University Press.
- Davidson, Donald. 1986. "A Coherence Theory of Truth and Knowledge." In LePore 1986.
- Hempel, Carl G. 1935. "On the Logical Positivists' Theory of Truth." *Analysis* 2, no. 4.
- Horwich, Paul. 1982. "Three Forms of Realism." *Synthese* 51: 181–201.
- LePore, Ernest, ed. 1986. *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. New York: Blackwell.
- Lynch, Michael. 1988. *Truth in Context*. Cambridge: MIT Press.
- Macdonald, Margaret, ed. 1954. *Philosophy and Analysis*. Oxford: Basil Blackwell.
- Plantinga, A. 1982. "How to Be an Anti-realist." *Proceedings and Addresses of the American Philosophical Association* 56: 47–70.
- Putnam, Hilary. 1978. *Meaning and the Moral Sciences*. London: Routledge and Kegan Paul.
- Putnam, Hilary. 1981. *Reason, Truth, and History*. Cambridge: Cambridge University Press.
- Rorty, Richard. 1979. *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.
- Rorty, Richard. 1982. *Consequences of Pragmatism*. Minneapolis: University of Minnesota Press.
- Williams, Michael. 1977. *Groundless Belief*. New Haven: Yale University Press.

---

## Contextual Semantics and Metaphysical Realism: Truth as Indirect Correspondence

Terence Horgan

Metaphysical realism is the view that there is a mind-independent, discourse-independent, world. Metaphysical antirealism is the view that there is no such world. Neither doctrine officially says anything about the nature of truth. Someone could espouse metaphysical realism, for example, and also espouse some version of *minimalism* about truth—roughly and generically, the view that the concept of truth is pretty much exhausted by instances of this schema:

The statement that *S* is true iff *S*.<sup>1</sup>

Likewise, someone could espouse metaphysical antirealism and also espouse some version of minimalism about truth.

Metaphysical realism, despite its official silence about the nature of truth, is often incorporated into a package-deal position that includes a correspondence conception of truth. Hilary Putnam describes this overall package (which he himself labels ‘metaphysical realism’) as a view that assumes the following:

1. A world consisting of a definite totality of discourse-independent objects and properties
2. ‘Strong bivalence’, i.e., that an object either determinately has or determinately lacks any property *P* that may significantly be predicated of that object
3. The correspondence theory of truth in a strong sense of ‘correspondence’, i.e., a predicate corresponds to a unique set of objects, and a statement corresponds to a unique state of affairs, involving the properties and objects mentioned in (1), and is true if that state of affairs obtains and false if it does not obtain<sup>2</sup>

Let *direct-correspondence metaphysical realism* (for short, DCMR) be the position that asserts each of (1) through (3). It is a combination of metaphysical realism properly so-called—embodied in thesis (1)—and a traditional conception of truth as direct correspondence to the mind-independent, discourse-independent, world—embodied in theses (2) and (3).

Contemporary versions of DCMR usually adopt a form of the correspondence theory of truth that I will call *referential semantics*. One key component of referential semantics is the contention that for natural languages (as well as for the formal languages typically studied by logicians), truth is characterizable via a Tarski-style recursive truth definition. Such a definition will employ semantic concepts like a singular term's *denoting* an object and a predicate's *applying to* an object. (Denotation and applicability may be conveniently lumped together under the blanket term 'reference'.) A second, even more crucial, component of referential semantics is the contention that reference should itself be understood in a robustly realist way, i.e., as involving genuine, direct, word-world relations. (So-called "causal theories of reference" constitute one sort of attempt to provide a robustly realist account of reference.)

Referential semantics might or might not be formulated so as to invoke the "states of affairs" mentioned in (3) above, or the properties mentioned in (1) and (3). Some advocates of DCMR are nominalists who deny the existence of properties but who still claim that there are definite referential relations linking our singular terms to definite discourse-independent things and linking our predicates either to definite classes whose members are definite discourse-independent things that "satisfy" the predicates, or at any rate (for nominalists who do not countenance classes) to those "satisfies" themselves. Furthermore, referential semantics yields a vigorously realist notion of truth as "correspondence" even without invoking states of affairs as entities to serve as correspondents for whole sentences. In addition, someone might hold that although there is a mind-independent, discourse-independent world as asserted by (1), nevertheless this world includes objects and/or properties that are essentially vague. And some who believe in such "ontological vagueness" also might deny that there is a *definite* totality of discourse-independent objects and properties, and/or deny that "strong bivalence" holds for vague objects and properties. So although claims (1)

to (3) as formulated by Putnam constitute a useful working characterization of the package-deal position I call DCMR, there are these caveats: (i) some versions of DCMR do not posit properties, (ii) some do not posit states of affairs, and (iii) some posit an “ontological vagueness” that perhaps does not obey strong bivalence and perhaps goes contrary to the idea of a definite totality of discourse-independent objects and properties.

In contrast to DCMR, contemporary versions of metaphysical anti-realism invariably reject referential semantics—as they must, since they do not acknowledge discourse-independent objects, properties, or states of affairs. Normally the antirealists repudiate, not Tarskian truth theory per se, but rather the conception of reference as a direct linkage between language and a discourse-independent, mind-independent, world. Anti-realist positions sometimes are wedded to a minimalist conception of truth, and sometimes to a radically epistemic conception of truth as “warranted assertibility” or some suitably idealized variant thereof. (I will call the latter conception *neopragmatist semantics*.)

In this paper I will propose and defend a position that combines metaphysical realism properly so called—essentially thesis (1) of DCMR—with a nontraditional, more liberal, version of the correspondence conception of truth that rejects theses (2) and (3) of DCMR. Correspondence, as I construe it, is very often an indirect relation between language (or thought) and denizens of an independently existing world; the traditional direct kind of correspondence presupposed by referential semantics is a limit case. I will call my position *indirect-correspondence metaphysical realism* (for short, ICMR). The core of this view is a general conception of language-world relations I call *contextual semantics*.

In section 1, I will describe the basic framework of contextual semantics, in a way that makes clear both (i) how the framework presupposes that there is a mind-independent, discourse-independent, world, and (ii) how the framework accommodates the idea that truth is often a matter of indirect, rather than direct, language/world correspondence. In section 2, I will argue the theoretical virtues of ICMR, as articulated within the framework of contextual semantics, over against DCMR. In section 3, I will defend ICMR on its other flank, arguing its theoretical virtues over against metaphysical antirealism. These first three sections describe and motivate the position I espouse.



In sections 4 and 5, I will extend the position in two ways. In section 4, I will briefly set forth and defend the generic approach to vagueness that I call “transvaluationism,” and I will explain why the transvaluationist approach to vagueness compels the conclusion that for all vague discourse, truth is indirect correspondence rather than direct correspondence. In section 5, I will sketch the nondescriptivist metaethical position described and defended in Timmons 1999 and Horgan and Timmons 2000, and I will explain how contextual semantics can accommodate truth ascriptions to moral statements even if, as Timmons and I contend, such statements bear neither direct nor indirect correspondence relations to the mind-independent, discourse-independent, world.

## 1 Contextual Semantics: The Core Framework

Contextual semantics has been evolving and developing in a series of writings.<sup>3</sup> Articulating and exploring it in further detail is a large-scale, long-term, research project of mine. The overall framework includes theses not only about truth and falsity *per se*, but also about meaning, ontology, thought, and knowledge. Contextual semantics, as I think of it, is intermediate between referential semantics (for short, *referentialism*), with its conception of truth as direct language-world correspondence, and neopragmatist semantics (for short, *neopragmatism*), with its radically epistemic construal of truth.

In articulating the distinctive claims of contextual semantics and for related expository purposes, throughout this paper I will borrow from Hilary Putnam the device of sometimes writing in small capital letters terms and phrases like ‘object’, ‘property’, and ‘the world’; this makes it unambiguously clear when I mean to be talking about denizens of the *mind-independent, discourse-independent, world*—the world whose existence is denied by metaphysical antirealists. (Antirealists typically regard as perfectly legitimate various everyday uses of the lowercase expressions, and some of their philosophical uses as well. The capitalization convention guarantees that claims that I intend to be incompatible with metaphysical antirealism will be construed as I intend them, rather than receiving a “compatibilist” reading.)

I will set forth the core framework of contextual semantics as a list of theses, interspersed with commentary.

- (1) Truth is *semantically correct assertibility*; falsity is *semantically correct deniability*.

Since we deny statements by asserting their negations, a statement is correctly deniable just in case its negation is correctly assertible. So henceforth I will usually speak only of “semantically correct assertibility” or “semantic correctness.” The relevant notion of semantic correctness has nothing to do with matters of etiquette. A statement can be semantically correct, in the relevant sense, even if it would be impolite, impolitic, or otherwise inappropriate to utter it.

- (2) Contrary to neopragmatism, truth is not radically epistemic, for semantically correct assertibility is distinct from warranted assertibility, and even from “ideal” warranted assertibility (Putnam 1981, 1983) and from “superassertibility” (Wright 1987, 1992).<sup>4</sup>

This thesis says, in effect, that the kind of *semantic* normativity that makes for truth and falsity is not reducible to epistemic normativity.

- (3) Semantic standards are not monolithic within a language. Instead, they vary somewhat from one context to another, depending upon the specific purposes our discourse is serving at the time.

Not only do semantic standards often vary from one mode of discourse to another, but they also often vary within a given mode of discourse. For instance, what counts as a *flat surface* is subject to contextually variable parameters within a given discourse. Similarly, what counts as the contextually eligible referent of a definite description like ‘that guy we were talking with awhile ago’, in a situation where several distinct entities in the relevant domain of quantification are eligible referents, is subject to contextually variable parameters. (Such parameters determine what David Lewis (1979) calls “the score in the language game.”)

- (4) Contrary to metaphysical antirealism, semantic correctness is normally a joint product of two factors: (i) the relevant assertibility standards and (ii) how things actually are in THE WORLD.

I call the operative semantic standards in a given discourse context *maximally strict* if they have this feature: under these standards a sen-

tence counts as correctly assertible (i.e., as true) only if there are OBJECTS and PROPERTIES in THE WORLD answering to each of the sentence's constituent singular terms, constituent assertoric existential quantifications, and constituent predicates.<sup>5</sup> The next two theses employ this notion.

- (5) Contrary to referentialism, our discourse often employs semantic standards that are not maximally strict.

That is, even though truth does typically depend upon how things are with THE WORLD, often this dependence is not a matter of direct correspondence between the constituents of a true sentence and OBJECTS and PROPERTIES. When the semantic standards are not maximally strict, the dependence is indirect.

Under contextual semantics, there is a whole spectrum of ways in which a statement's correct assertibility can depend upon THE WORLD.<sup>6</sup> At one end of the spectrum are statements governed by semantic standards, in a given context of usage, that are maximally strict (and thus coincide with those laid down by referentialism). Under these standards, a statement is true only if some unique constituent of THE WORLD answers to each of its singular terms, and at least one such entity answers to each of its assertoric existential-quantifier expressions. (Statements asserted in order to make serious ontological claims—like 'There exists a God', as asserted by a conventional theist—presumably are governed by maximally strict semantic standards.) At the other end of the spectrum are statements whose governing semantic standards, in a given context, *alone* sanction those statements as semantically correct, independently of how things are with THE WORLD. (Statements of pure mathematics are plausible candidates for this status.) Both ends of the spectrum are limit cases, however. Various intermediate positions are occupied by statements whose semantic correctness, in a given context, does depend in part on how things are with THE WORLD, but where this dependence does not consist in direct correspondence between the referential apparatus of the statements (its singular terms, quantifiers, and predicates) and OBJECTS or PROPERTIES in THE WORLD.

As a plausible example of a statement that normally would be governed by semantic standards falling at an intermediate point in the spectrum just described, consider (a):

- (a) Beethoven's fifth symphony has four movements.

The correct assertibility of (a) probably does not require that there be some ENTITY answering to the term 'Beethoven's fifth symphony' and also answering to the predicate 'has four movements'. Rather, under the operative semantic standards, (a) probably is semantically correct (i.e., true) by virtue of other, more indirect, connections between the sentence and THE WORLD. Especially germane is the behavior by Beethoven that we could call "composing his fifth symphony." But a considerably wider range of goings-on is relevant too: in particular, Beethoven's earlier behavior in virtue of which his later behavior counts as composing his *fifth* symphony, and also a broad range of human practices (including the use of handwritten or printed scores to guide orchestral performances) in virtue of which such behavior by Beethoven counts as "composing a symphony" in the first place. Further plausible examples of statements governed by semantic standards that are not maximally strict include the following:

- (b) The University of Memphis is a public institution.  
 (c) Mozart composed 27 piano concertos.  
 (d) There are more than 20 regulatory agencies in the U.S. Federal Government.  
 (e) Quine's *Word and Object* is an influential book.

Although contextual semantics asserts that the operative semantic standards governing truth (semantic correctness) can vary from one context to another, it also asserts that contextually operative *meta-linguistic* semantic standards normally require truth ascriptions to obey Tarski's schema T:

- (6) Even in discourse contexts where the operative semantic standards are not maximally strict, these standards typically sanction as true (i.e., as semantically correct) instances of Tarski's equivalence schema:  
 (T) "*p*" is true if and only if *p*.<sup>7</sup>

Thesis (6) says, in effect, that the contextually operative semantic standards governing the truth predicate normally operate "in tandem" with

those governing first-order discourse; as I put it in Horgan 1986b, truth talk is *assertorically consistent* with first-order talk.

If contextual semantics is right, so that truth is intimately bound up with semantic standards, then meaning too is intimately bound up with these standards.<sup>8</sup> Intuitively and pretheoretically, meaning is what combines with how THE WORLD is to yield truth. Thus, if truth is correct assertibility under operative semantic standards, then the role of meaning is played by *the semantic standards themselves*. So matters of meaning are, at least in large part, matters of operative semantic standards.<sup>9</sup> Contextual semantics makes the following nonreductionist claim about matters of meaning:

- (7) In general, if a statement *S* is semantically correct under certain frequently operative semantic standards but *S* is not semantically correct under maximally strict semantic standards, then *S* is not equivalent in meaning to—or approximately equivalent in meaning to, or “intensionally isomorphic” to, or “regimentable” into—a statement that is correctly assertible under maximally strict semantic standards.

Thesis (7) is one I came to believe after pursuing for some time the project of trying to systematically paraphrase (“regiment” in Quine’s terminology) statements whose surface grammar embodies an apparent commitment to ontologically dubious entities into a more austere idiom that eschews reference to such entities. Although the paraphrase strategy can sometimes be carried through piecemeal for certain local segments of discourse, very often it evidently will *not* work. (Trying to implement the strategy for statements like (a) through (e) caused me to lose faith in it.)

Under contextual semantics the issue of ontological commitment becomes much more subtle than it is under referential semantics, because whenever the contextually operative semantic standards are not maximally strict, the so-called “referential apparatus” of our discourse need not connect directly to OBJECTS and PROPERTIES in the world in order for our statements to be true. Here then are several theses concerning ontology:

- (8) It is necessary to distinguish between *regional ontology*, which concerns the range of putative entities overtly posited by a given mode of discourse, and *ultimate ontology*, which concerns the

range of entities posited by statements that are correctly assertible under maximally strict semantic standards.

- (9) Quine's well-known criteria of "ontological commitment" are directly relevant only to regional ontology, not to ultimate ontology.
- (10) Determining the ultimate ontological commitments of our scientific and nonscientific discourse is a methodologically subtle matter in which we inquire what THE WORLD is like IN ITSELF in order to be correctly describable, under various contextually operative semantic standards, by those statements that are true in everyday life and in science.

Whatever exactly the right story is about ultimate ontology, it seems quite plausible that a complete and accurate accounting of what there really is in THE WORLD need not include entities like the State of Tennessee, the U.S. Federal Government, Mozart's 27th piano concerto, or Quine's book *Word and Object*. In terms of ultimate ontology, such entities are artifacts of our conceptual scheme; they are not mind-independently, discourse-independently REAL. Although THE WORLD does normally contribute to the truth or falsity of statements that are regionally ontologically committed to such entities, it does so quite indirectly.

Although contextual semantics rejects the epistemic reductionism of neopragmatism, it also acknowledges something importantly right that is reflected in that approach:

- (11) Contextually operative semantic standards are typically intimately linked to prototypical evidential conditions for statements.

We all know quite well, for instance, what sorts of evidence are relevant to claims like (a) through (e), and the kind of evidence we would look for has rather little to do with the philosophical question of whether ultimate ontology should include entities like SYMPHONIES, PIANO CONCERTOS, BOOKS, or a FEDERAL GOVERNMENT. Under the semantic standards operative in ordinary-discourse contexts, it is quite appropriate that the relevant epistemic standards should bypass the issue of ultimate ontology, for the semantic standards themselves are not maximally strict. There is a comparatively small "conceptual gap" between the epistemic standards

for warranted assertibility and the semantic standards for *correct* assertibility (even though semantic standards are not reducible to epistemic ones). There is a gap, though. For instance, Mozart might have engaged in behavior that would count as composing a 28th piano concerto, even though there is no extant evidence available. If so, then the statement that Mozart composed 27 piano concertos would be both warrantably assertible and yet false. In part, the gap between warranted assertibility and semantic correctness results from the holistic aspects of evidence:

- (12) Our attributions of truth and falsity usually are defeasible even under prototypical evidential conditions, for the semantic correctness of any given statement normally depends, in part, on the semantic correctness of various other statements that are assumed, in a given evidential situation, to be semantically correct.

As Quine and Duhem stressed long ago, our statements really face the tribunal of empirical evidence jointly, not singly.

Contextual semantics also includes a psychological dimension (which, as John Biro has urged on me, might better be called *psychosocial*):

- (13) Which semantic standards are the operative ones, in any given context of discourse, depends largely on the contextually attuned, socially coordinated, truth-judging and falsity-judging dispositions of competent speakers.

The interconnections between the judgment dispositions of competent speakers and the contextually operative semantic standards are typically fairly subtle; surely no crudely reductive account will work. (For one thing, even competent speakers often exhibit linguistic-performance errors. For another, a competent speaker's judgment dispositions are normally more directly indicative of what is *warrantably* assertible from the available evidence, and sometimes this diverges from what is semantically correct under contextually operative semantic standards.) Nonetheless, such socially coordinated psychological dispositions do figure importantly in determining the contextually operative semantic standards.

Contextual semantics has various points of contact with the views of other philosophers on language-world relations. It seems to me to be a natural and plausible extension, for instance, of the treatment of con-

textually variable discourse parameters in Lewis 1979. Likewise, it seems to me to be a natural further step in a direction already taken by advocates of philosophical projects of “regimentation,” namely, the direction of denying that the surface ontological commitments of true statements always constitute *ultimate* ontological commitments. I have already mentioned that it accommodates certain motivating ideas in neopragmatism (and in verificationism), but without the mistake of embracing epistemic reductionism. There are echoes of Carnap’s famous contention that a “linguistic framework” can automatically sanction existence claims concerning the entities posited by the framework, and that such existence claims are ontologically innocent (Carnap 1950). The approach is somewhat similar to the treatment of truth and ontology in Sellars 1963, 1968.<sup>10</sup> Finally, contextual semantics seems to me rather similar in spirit to the general approach to truth, and to philosophical debates about realism and antirealism concerning various forms of discourse, in Wright 1992.<sup>11</sup>

## 2 Indirect-Correspondence versus Direct-Correspondence Metaphysical Realism

ICMR, in the version I espouse, incorporates the core framework of contextual semantics. I will briefly argue for ICMR by describing some of its principal theoretical advantages over its main rivals: DCMR on the one hand and metaphysical antirealism on the other hand. In this section I focus on the comparison to DCMR.

One important theoretical advantage of ICMR over DCMR is ontological. In theorizing about ultimate ontology, as in theory construction more generally, parsimony is desirable. Contextual semantics makes possible a substantial paring down of the ultimate ontological commitments of our discourse. The desirable simplifications come not merely from minimizing the number of distinct kinds of entities one posits, but also from avoiding a baroque relational network involving entities at various ontological “levels”: some in space-time (e.g., tables, electrons) and others not (e.g., numbers, attributes); some (e.g., corporations, nations) synchronically supervenient upon others (e.g., persons, buildings, land masses); some causally related to one another and others not; etc.



But although parsimony is highly desirable in ultimate ontology, there is little reason to believe either that putative higher-level entities can be reductively identified with entities posited by a sparse, uniform, materialist ontology or that talk about the “irreducible” entities can be systematically paraphrased (“regimented,” as Quine puts it) into a more austere idiom that avoids Quinean ontological commitment to the offending entities.<sup>12</sup> Hence the attractiveness of contextual semantics for a metaphysical realist over the referential semantics of DCMR. For the former can, while the latter evidently cannot, allow for genuine truth without ultimate ontological commitment even when reductive identifications are not forthcoming and even when the true statements about higher-level entities cannot be “paraphrased away.”

A second advantage of ICMR over DCMR concerns knowledge. Contextual semantics holds out the promise of a tractable epistemology, particularly within the framework of a naturalistic conception of human beings as complex physicochemical systems. Consider, for instance, our knowledge of statement (a) above. Philosophers who are concerned about ontological questions in aesthetics and who approach these questions within the framework of standard referential semantics have been much exercised by the ontological status of musical works of art. If we begin by assuming that a sentence like (a) cannot be true unless some OBJECT answers to the term ‘Beethoven’s fifth symphony’, then we are apt to suppose (as many have) that this term denotes a complex, internally structured *universal*—an abstract entity that exists eternally and is not part of the spatiotemporal causal nexus. But once we suppose this, it becomes very hard to see how mere humans could ever have *knowledge* about symphonies, such as the knowledge expressed by (a). For we cannot come into any sort of causal contact with these putative entities, but rather can only causally interact with those concrete things we call “performances of Beethoven’s fifth symphony,” “copies of the score of Beethoven’s fifth symphony,” etc. One is tempted to say, of course, that we know the symphony itself via our knowledge of concreta that “token” it or “describe” it. But this only pushes the epistemological problem back a step. For how could we know that a given event tokens a certain abstract OBJECT, or that a given manuscript describes it, unless we could somehow directly compare the event or the manuscript with the

abstract OBJECT itself? Yet it is most implausible—especially on a naturalistic conception of human beings—to suppose that we can have some sort of quasi-perceptual cognitive communion with an entity that has no spatiotemporal location and does not causally interact with anything. And if someone replies by saying that our performance instances and score copies are instances of, or copies of, a particular abstract universal *because we stipulate them to be*, rather than because we somehow directly compare those concrete things with the putative abstract entity, then again the epistemological puzzles are merely pushed back. For now it becomes very hard to see either (i) how we can justifiably claim to know that there really exist such putative abstract OBJECTS as symphony types at all or (ii) how our stipulative acts could link a term like ‘Beethoven’s fifth symphony’ to one such specific complex universal, rather than to some other one that is isomorphic to the first.

These highly vexing epistemological problems do not arise under contextual semantics, since the semantic standards for symphony talk are standards under which a sentence like (a) can be semantically correct even though no actual denizen of THE WORLD answers to the term ‘Beethoven’s fifth symphony’. Similarly for a wide variety of other kinds of discourse. (Talk about numbers, sets, and other mathematical entities is another especially salient example.) Thus ICMR has a greater potential for rendering our knowledge claims justifiable than does DCMR. For, ICMR incorporates contextual semantics, whereas DCMR is constrained to incorporate either referential semantics or some other conception of truth as direct language-WORLD correspondence.

### 3 Indirect-Correspondence Metaphysical Realism versus Metaphysical Antirealism

How is ICMR, based upon contextual semantics, preferable to metaphysical antirealism (MAR)? I will argue that the fundamental and all-important difference between the two positions is that ICMR is intelligible, whereas MAR is not.

DCMR, including the conception of language-world relations embodied in referential semantics, is in many ways the naively natural view, the view seemingly embodied in our everyday thought and discourse. We

believe that many objects in our environment exist quite independently of ourselves, our mental activity, and our discourse: trees exist unperceived; stars and galaxies existed before humans ever came on the scene; many of the subatomic particles out of which our own bodies are composed have existed since the Big Bang; etc. We believe, likewise, that many of our singular terms determinately denote these objects, and many of our predicates determinately apply to them: expressions like ‘the dogwood tree in Terry Horgan’s back yard’, ‘the Andromeda galaxy’, ‘is a star’, and ‘is a neutrino’ stand in quite definite reference relations to these entities.

No plausible philosophical position can simply deny outright that there is a mind-independent, discourse-independent world, or that our words stand in determinate reference relations to objects in this world. For at the level of ordinary discourse, these claims are virtual platitudes. Rather, any philosophical position that backs away from DCMR and referential semantics must accommodate these platitudes, even if the position maintains that they are seriously mistaken when invoked, outside of ordinary contexts, as answers to certain philosophical and/or theoretical questions, such as philosophers’ questions about ontology.

This methodological constraint creates a dilemma for the philosopher who seeks to reject DCMR and referential semantics. One needs to employ ordinary language to convey one’s position, even though one might spice it up with specialized jargon or other related linguistic devices (such as my use of words in small capital letters in this paper), and yet one also needs to use language in a somewhat nonstandard way to explain why one repudiates the platitudes as guides to philosophers’ questions. I will call this conundrum the *Kantian dilemma*, because Kant faced it so vividly and explicitly in trying to articulate his own philosophical position. He acknowledged the platitudes, by allowing that trees, stars, and the rest are “empirically real.” And yet he also maintained that these objects are, in a somewhat special and philosophical sense, mind-dependent: they are “transcendentally ideal.” The Kantian dilemma poses a serious *prima facie* problem of intelligibility: since language must inevitably be strained in articulating a philosophical position distinct from DCMR, there is a constant danger of lapsing into incoherence.

So if one wants to claim that many or all of the entities we talk about are, in some philosophically special sense, “mind-dependent” and “discourse-dependent,” then one takes on a heavy burden of making one’s position intelligible. This burden includes explaining cogently what the philosophical notion of mind-dependence and discourse-dependence comes to, motivating the claim that the objects of ordinary and scientific discourse have this status, and reconciling this claim with the platitudinous fact that in the ordinary sense of ‘depend’, many of those objects do *not* depend upon mind or language. MAR supporters and I both face this task. The defenders of MAR tend not to acknowledge the burden, however. Instead, they often write as though the metaphysical realists are the only philosophers who take language on holiday—a curious stance in view of how initially peculiar-looking is the claim that the things like trees, stars, and electrons are mind-dependent and discourse-dependent!

Since advocates of MAR repudiate THE WORLD entirely, they face an especially virulent form of the Kantian dilemma. Not only must they explain the special philosophical sense of mind/language dependence; they must also explain how there could possibly be entities that depend—in the relevant sense of ‘depend’—on mind and discourse even though there allegedly does not exist anything that does not depend on them (in that special sense). Yet *prima facie*, this task looks impossible. To see the problem, consider Putnam’s own succinct summary of his own antirealist view:

In short, I shall advance a view in which the mind does not simply ‘copy’ a world which admits of description by One True Theory. But my view is not a view in which the mind *makes up* the world, either (or makes it up subject to constraints imposed by ‘methodological canons’ and mind-independent ‘sense-data’). If one must use metaphorical language, then let the metaphor be this: the mind and the world jointly make up the mind and the world. (Or, to make the metaphor even more Hegelian, the Universe makes up the Universe—with minds—collectively—playing a special role in the making up.) (Putnam 1981, xi)

I do not object in principle to the use of metaphorical language in explaining a position that repudiates metaphysical realism; metaphor may be hard to avoid because of the internal logic of such a position. But Putnam’s description of his position seems unintelligible even at the metaphorical level, because the metaphors only pose again the question of how anything mind-dependent could ever exist unless something

mind-independent also existed. How could mind and the world create *themselves*, if both are genuinely mind-dependent? Wouldn't a mind-dependent Universe require for its existence an already existing, meta-physically real MIND, such as a Berkeleyan MIND or a divine MIND?

These questions indicate how far Putnam is from providing a viable solution to the Kantian dilemma, and how heavy is the burden of proof that the dilemma has any solution within the framework of MAR. And other current versions of MAR seem no better off in this respect. So unless and until that burden is discharged by Putnam or some other champion of MAR, I think the reasonable course is to conclude, at least provisionally, that MAR is incoherent.

Of course, ICMR too must address the Kantian dilemma. But the prospects for successfully doing so, within the framework of contextual semantics, look fairly promising. Since THE WORLD plays an integral role within contextual semantics, ICMR's form of the Kantian dilemma does not include the impossible-looking task of explaining how absolutely *everything* could be mind-dependent.

Platitudes that the objects in our environment are mind-independent and discourse-independent, for instance, are accommodated thus: these statements are indeed literally true, under the semantic standards that actually prevail in contexts of discourse where such statements usually occur. The same goes for semantic platitudes like "The predicate 'is a cat' applies to cats," for metalinguistic discourse too is subject to context-dependent semantic standards (see Horgan 1986b). The special, philosophical sense in which many objects and properties are mind-dependent and discourse-dependent, on the other hand, comes essentially to this: (i) virtually any statement with regional ontological commitments fails to be true under the referentially strict semantic standards described by referential semantics, and (ii) although the truth or falsity of such statements, under various contextually operative semantic standards, is normally determined by how things are with THE WORLD itself, the standards do not actually require that THE WORLD contain ENTITIES of the kind to which the statements are regionally ontologically committed.

To what extent, one might well ask, do the objects and properties to which ordinary and scientific discourse is regionally ontologically committed turn out to be mind-dependent and discourse-dependent in this

philosophically special way? To a great extent indeed, say I. One important route to this conclusion is by way of the phenomenon of vagueness, a topic to which I now turn.

#### 4 The Transvaluationist Conception of Vagueness and Its Consequences

Vagueness is ubiquitous in language and thought, both in common sense and in science. Many of the objects and properties posited in common sense and in science are vague. For example, many posited objects are vague with respect to their spatiotemporal boundaries or vague with respect to their synchronic composition, and many posited properties are vague with respect to their range of instantiation. When one attends carefully to the nature of vagueness, some striking implications emerge, namely, that vague OBJECTS and PROPERTIES are logically impossible, and hence that truth for vague discourse must be indirect correspondence rather than direct correspondence. Here I will summarize briefly the reasoning leading to these conclusions.<sup>13</sup>

An essential attribute of genuine vagueness is what Mark Sainsbury (1990) calls *boundarylessness*, a feature that can be characterized by reference to sorites sequences associated with vague terms. Consider a vague term, say 'heap', and consider a sorites sequence involving the given term in which the initial statement is true and the final statement is false, say a series of statements successively predicating the vague term 'heap' first of a pile of sand with 1 billion grains, then of an object produced by removing just one grain, then of an object produced by removing yet another single grain, and so forth down to a statement predicating 'heap' of a single grain of sand. To say that vagueness is boundarylessness is to say that in such a sequence, (i) initially there are true statements (with each predecessor of any true statement being true), (ii) later there are false statements (with each successor of a false statement being false), and (iii) there is no determinate fact of the matter about the transition from true statements to false ones. Condition (iii) requires not only that there be no determinate *abrupt* transition from true statements to false ones, but also that the truth/falsity transition should involve no determinate semantic transitions at all. Thus it also precludes, for example,

an overall true-to-false transition involving first a determinate abrupt transition from truth to the semantic status “indeterminate whether true or false,” and later another determinate abrupt transition from this in-between status to falsehood.<sup>14</sup>

If one considers what it would take to fully accommodate boundarylessness—that is, accommodate it in a way that thoroughly eschews arbitrary semantic transitions of any kind—one finds that, for the successive statements in a sorites sequence, there are semantic requirements in play that cannot be simultaneously satisfied. Boundarylessness has two conceptual poles. On one hand, there is an individualistic pole, applicable to individual pairs of adjacent statements in a sorites sequence. Namely, for any pair of adjacent statements, the two statements must have the same semantic status (truth, falsity, indeterminateness, or whatever). Otherwise, the juncture of the two statements would constitute a determinate semantic transition point in the sequence, contrary to the claim that there is no determinate fact of the matter about semantic transitions in a sorites sequence. On the other hand, there is also a collectivistic pole in the notion of boundarylessness, applicable globally with respect to a sorites sequence as a whole. Two collectivistic requirements apply. First, it is impermissible to indefinitely iterate the individualistic-pole requirement for successive adjacent pairs of statements, in the manner of paradoxical sorites arguments. Second, there is simply no determinate collective assignment of semantic status to all the statements in a sorites sequence. These individualistic and collectivistic requirements cannot be jointly satisfied, for the only way in which a sorites sequence could fully conform to the individualistic pole would be for every statement in the sequence to have the same semantic status. (This is the lesson of the sorites paradox, which emanates directly from the individualistic pole of boundarylessness.) So boundarylessness is logically incoherent in a specific way: it imposes mutually unsatisfiable semantic standards upon vague discourse.

The specific kind of logical incoherence exhibited by vagueness needs to be distinguished from a stronger, and highly malevolent, kind of logical incoherence. Vagueness does involve *weak* logical incoherence, namely, the presence of mutually unsatisfiable semantic standards governing vague discourse (and vague thought content). But this does not

necessarily bring in its wake *strong* logical incoherence, namely, commitment to individual statements that are logically contradictory, such as statements of the form  $p \ \& \ \neg p$ .<sup>15</sup>

The semantic standards that govern vague discourse can, and do, incorporate weak logical incoherence without the strong kind. How? Briefly, the story goes as follows. Insofar as vague discourse exhibits boundarylessness, incompatible individualistic and collectivistic semantic requirements are indeed in force. That is, no requirement is defeated by any others, in the sense of having defeasibility conditions that are satisfied by the presence of the competing and incompatible requirements. But these competing requirements are not on a par with one another either. The collectivistic-pole requirements *dominate* the individualistic-pole requirements without defeating them; that is, to the extent that the requirements conflict, truth is determined by the collectivistic-pole requirements. In practice, this means that paradoxical sorites arguments are to be eschewed. It also means that one must not acknowledge the existence of any determinate semantic transitions (even unknown or unknowable ones) in a sorites sequence. (Semantic status still must conform partially to individualistic-pole requirements, however. For instance, it is never the case, for any specific pair of adjacent statements in a sorites sequence, that the two statements differ in semantic status.<sup>16</sup>) So the semantic standards governing vague discourse are *logically disciplined*, in virtue of the dominance of collectivistic-pole requirements, though without the defeat of the individualistic-pole requirements. Because of this logical discipline, no logically incoherent statement is true under those standards; strong logical incoherence is avoided.

*Transvaluationism* is my name for the general approach to vagueness I have been describing. Transvaluationism claims that vagueness is weakly logically incoherent without being strongly logically incoherent. It also claims that vagueness is viable, legitimate, and indeed essential in human language and thought; its weak logical incoherence is benign rather than malevolent. Just as Nietzsche held that one can overcome nihilism by embracing what he called the transvaluation of all values, transvaluationism asserts that vagueness, although logically incoherent in a certain way, can and should be affirmed and embraced, not nihilistically repudiated.<sup>17</sup>



If vagueness is really boundarylessness, as it certainly appears to be, then since boundarylessness involves disciplined weak logical incoherence, an adequate treatment of vagueness will have to be some version of transvaluationism. Moreover, transvaluationism is a fairly generic approach, potentially open to further development and articulation in a variety of different ways. Thus numerous details about the logic and semantics of vagueness remain open within the generic conception and might get handled differently in different versions.<sup>18</sup> But regardless of how the details go, any account of vagueness that seriously comes to grips with boundarylessness must be a version of transvaluationism, whether its proponents acknowledge this fact or not.<sup>19</sup> In effect, specific proposals amount to suggested strategies for implementing the dominance of the collectivistic semantic standards over the individualistic ones without the defeat of the latter.<sup>20</sup>

We are ready now to draw out the powerful implications of boundarylessness for metaphysics and for semantics. First, for metaphysics. THE WORLD cannot be logically incoherent, even in the weak way: it cannot have features that are the ontological analogues of mutually unsatisfiable semantic standards. For example, there cannot be a genuine PROPERTY H (for 'heaphood') and a sequence of sand conglomerations each of which has one fewer grain than its predecessor such that (i) initially in the sequence there are instances of H (with each predecessor of an H instance being an H instance), (ii) eventually there are non-H instances (with each successor of a non-H instance being a non-H instance), and (iii) for each pair of successive piles in the sequence, either both are H instances or both are non-H instances or both are neither. For, the only way to satisfy condition (iii) would be for all the piles to have the same status vis-à-vis H. But vagueness involves boundarylessness essentially, and boundarylessness involves weak logical incoherence essentially. Hence there cannot be ontological vagueness, and in particular, there cannot be vague OBJECTS or vague PROPERTIES.

Next, for semantics. Weak logical incoherence is a feature of the contextually operative semantic standards governing vague discourse in ordinary contexts of usage. Hence truth for discourse involving vagueness cannot be a matter of direct language-WORLD correspondence, for this would mean that THE WORLD itself would have to exhibit the same

logical incoherence that is present in vagueness, and this is impossible. Thus—barring the wildly implausible, nihilistically self-defeating, position that vague statements are never true—truth for vague discourse must be a form of indirect correspondence. Furthermore, if this is so, then there is no particular problem about the weak logical incoherence of the operative semantic standards as long as these standards are logically disciplined, and hence are not also strongly logically incoherent.

The upshot of these considerations is that the only viable general approach to vagueness is one that conceives it nonontologically (thereby repudiates all vague OBJECTS and PROPERTIES) and construes truth (for vague discourse) as indirect correspondence. The correct ultimate ontology, whatever it turns out to be, cannot admit vague OBJECTS or PROPERTIES, and an appropriate semantics for discourse employing vague posits will have to treat truth for such discourse as indirect correspondence. These conclusions have very wide application indeed, since vastly many of the posits employed both in common sense and in science are vague.<sup>21</sup>

Within contextual semantics, indirect correspondence is best understood as semantic correctness under contextually operative semantic standards that are not referentially strict. As I said earlier, we sometimes employ language under limit-case, referentially strict, direct-correspondence semantic standards. These are the standards appropriate for serious ontological inquiry. When they are in play, so is classical two-valued logic. Under this limit-case use of language, sorites reasoning can be correctly employed to construct *reductio ad absurdum* arguments against the existence of vague PROPERTIES or vague OBJECTS, including, of course, not only MOUNTAINS and CLOUDS but also TABLES, CHAIRS, and (regrettably) PERSONS.<sup>22</sup> Fortunately, however, ordinary uses of vague language in everyday contexts and even in science are not governed by limit-case semantic standards, and normally the contextually operative standards conspire with nonvague REALITY to render much of our vague discourse true.

## 5 Nondescriptivist Cognitivism in Metaethics

According to the core framework of contextual semantics, truth is semantically correct assertibility under contextually operative semantic

standards, and falsity is semantically correct deniability under such standards. Truth (and falsity) are a joint product of the contextually operative semantic standards on the one hand, and how things are with THE WORLD on the other hand. Truth is thus *correspondence* to THE WORLD—typically *indirect* correspondence, since normally the contextually operative semantic standards are not referentially strict. (Likewise, falsity is *noncorrespondence* to THE WORLD—typically indirect noncorrespondence.) Mark Timmons and I call the contextually operative semantic standards for a given discourse *tight*, vis-à-vis certain statements within the discourse, if these standards do indeed conspire with THE WORLD to render those statements correctly assertible or correctly deniable (Horgan 1995a, 1996; Timmons 1999). The overall assertoric content of such statements is *descriptive* content, i.e., “a-way-the-world-might-be” content.

An important extension of the core framework arises if one claims that for certain kinds of discourse, speakers employ declarative sentences to make genuine statements with full-fledged cognitive content (i.e., content that is belief-eligible and assertoric), even though the contextually operative semantic standards applicable to these statements are not tight. Thus, the statements would not stand in relations of correspondence (or noncorrespondence) to THE WORLD, not even in relations of *indirect* correspondence (or noncorrespondence).

Timmons (1999) and Horgan and Timmons (2000) argue that *moral* statements have the features just described: although they have full-fledged cognitive content, typically they are governed by contextually operative semantic standards that are not tight; hence, moral statements lack any objectively determinate semantic correct-assertibility or correct-deniability status. Instead, moral assertions are made from within a morally engaged stance reflective of the speaker’s own moral commitments. A moral assertion is thus a *stance-taking* speech act, an act through which (i) one expresses a specific moral commitment and (ii) one positions oneself, within the context of sociolinguistic dynamics, with respect to the particular issue at hand. It is an action-guiding speech act, whose typical role within interpersonal dynamics involves reasons for action and being prepared to provide them. Moral statements are normally asserted *categorically*, over against competing moral stances that

might be occupied by those with whom one is communicating. Among the features that are semantically central to such categorical assertions are the following. First, moral statements typically effect a *demand*, of those to whom the statement is addressed, to behave in ways conforming to the moral stance being asserted. Second, they typically signal the speaker's readiness to back the moral stance being asserted with *non-subjective reasons*, i.e., reasons that do not appeal to matters of individual taste, personal preference, or the like. Third, they typically *challenge* the listener, insofar as he adopts some conflicting moral stance, to provide nonsubjective reasons in support of that stance and against the speaker's own stance.

The position I have been briefly sketching is what Timmons and I call *nondescriptivist cognitivism*. I lack the space here either to describe it more thoroughly or to summarize our arguments for it. But let me briefly address this question: what should one say about truth ascription to moral statements from the viewpoint of this nondescriptivist cognitivism about moral content and the wider theoretical framework of contextual semantics? Recall that according to contextual semantics, semantic standards frequently involve contextually variable parameters. Contextual variability can manifest itself not only among different kinds of discourse involving different subject matters but also within a single kind of discourse. Specifically, this can happen with respect to the standards governing the contextually appropriate uses of 'true' and 'false' as applied to moral statements. The flexibility of the parameters manifests itself by there being two different perspectives from which to appraise moral statements semantically.

From a *morally detached* perspective, one uses 'true' and 'false' in such a way that a statement counts as true (false) just in case it is semantically correctly assertible (correctly deniable) solely by virtue of the contextually operative semantic standards plus THE WORLD. One asks, in effect, whether or not the semantic standards governing the discourse conspire with features of THE WORLD to yield semantically correct assertibility (or deniability) of statements in the discourse; thus, 'true' and 'false' are being used to express correspondence or noncorrespondence (either direct or indirect) between language and THE WORLD, and they are only applicable to statements governed by tight semantic standards. Given the

claims lately made about moral discourse, when one judges from a morally detached perspective and thus simply in light of semantic standards, moral statements are neither correctly assertible nor correctly deniable, and so they are neither true nor false. In other words, under the contextual parameters operative in judging from a detached perspective, the truth predicate does not properly apply to moral statements.

However, people can and do judge the truth of moral statements from within a *morally engaged* perspective, and in these contexts the use of 'true' runs in tandem with the object-level discourse. When speaking and judging metalinguistically in a morally engaged way, one's use of 'true' and 'false' vis-à-vis moral statements becomes morally assertoric itself, so that truth ascriptions are a fusion of semantic and moral evaluation. That is, the contextually operative parameters on the use of 'true' require that (1) one's semantically appropriate use of moral statements be reflective of one's own moral commitments and (2) the semantically appropriate use of truth talk for moral statements accord with schema T.

So, in answer to questions about moral truth, the view being proposed here is, in short, this. From a certain morally detached perspective—the perspective of theoretical inquiry—there is no moral truth (or falsity), since semantic standards alone do not conspire with THE WORLD to yield correct assertibility or deniability status to moral statements. Truth is correspondence to THE WORLD (either direct or indirect), and moral statements, being nondescriptive in their assertoric content, cannot enter into relations of language/WORLD correspondence (either direct or indirect). But from an engaged perspective, 'true' as predicated of moral statements results in a metalinguistic claim that is a fused semantic/moral assertion rather than a detached semantic assertion. Both ways of employing the truth predicate are legitimate in context because of the contextual variability of the semantic standards governing the notion of truth. One can properly use truth talk either way, but not in the same breath.

## Notes

This paper excerpts material from several prior papers of mine, especially Horgan 1991, 1995b, 1998b, 2000.

1. Here and throughout, I use the term 'statement' in a way that is conveniently ambiguous among various candidate truth bearers such as sentence tokens, sentence types, and propositions. The issues I address in this paper are largely orthogonal to debates about the nature of truth bearers. Also, although I will often speak of correspondence as a "language-world relation," I intend the use of the word 'language' within this locution to be similarly neutral about the nature of truth bearers.
2. Putnam 1983, 272. Putnam attributes to Michael Dummett the same three-part characterization of metaphysical realism.
3. Horgan 1986a, 1986b, 1991, 1994a, 1994b, 1995a, 1995b, 1996, 1998a; Horgan and Potrc, 2000; Horgan and Timmons 1993; Timmons 1999. Originally I called the view "language-game semantics" and later called it "psychological semantics," before Timmons and I decided on the current name.
4. I argue against epistemically reductionist construals of truth like those of Putnam (1981, 1983) and Wright (1987, 1992) in Horgan 1991, 1995a, 1996. One line of argument I use appeals to a brain-in-vat scenario in which the brain was originally embodied and has only recently, unwittingly, become envatted.
5. Nominalism, as an ontological position about properties, is something I will pass over in this paper to avoid complicating my discussion unnecessarily.
6. The metaphor of a spectrum is really too simple and unidimensional, but it serves my present expository purposes.
7. This leaves it open whether or not contextually operative assertibility standards typically sanction as true *all* instances of schema (T). In connection with vagueness, doubts can be raised about instances of (T) in which the statement replacing '*p*' is a vague predication involving a borderline case (e.g., a statement predicating 'bald' of someone who is a borderline case of baldness). Vagueness-related doubts can also be raised about instances of (T) in which '*p*' is replaced by certain quantificational statements (e.g., the statement 'For any *n*, if a person with *n* hairs on his head is bald, then a person with *n* + 1 hairs on his head is bald'). More on vagueness in section 4 below.
8. Contextual semantics, as it has so far been worked out, focuses more on truth than on meaning.
9. Variations in the operative semantic standards from one context to another generally do not, however, constitute differences in meaning. It is more accurate to view matters of meaning in the following way. (1) Generic semantic standards have certain contextually variable parameters. (2) Specific, contextually operative, semantic standards involve particular values of those parameters; these parameter values determine the current "score in the language game." (3) The generic semantic standards hold transcontextually, whereas the specific parameter values differ from one context to another. (4) Meaning remains constant transcontextually, because of the constancy of the generic semantic standards. (5) Contextual variability in parameter values constitutes a more subtle, more fine-grained, kind of semantic change than does change in meaning. As one might

put it, changes in parameter values yield a *différance*—not a difference—in meaning. (Moreover, as Bill Throop has pointed out to me, the term ‘meaning’ itself is evidently governed by semantic standards with contextually variable parameters: although the term is frequently used in the coarse-grained way just described, we do sometimes use the phrase ‘change in meaning’ to track more fine-grained semantic differences.)

10. See especially the essays “Truth and ‘Correspondence,’” “Grammar and Existence: A Preface to Ontology,” and “Some Reflections on Language Games” in Sellars 1963, and chapter 4 of Sellars 1968.

11. An important difference between Wright and me is that I vigorously eschew epistemic reductionism, whereas Wright (1992) remains officially neutral about it. Furthermore, his book can be read as supportive of the contention that truth, in any discourse, is the epistemically characterizable attribute he calls super-assertibility. In Horgan 1995a, 1996, I applaud Wright’s generic position but argue against an epistemically reductionist version of it.

12. In Horgan 1986b, I provide reasons to be skeptical about the prospects for systematic reductive identifications or systematic eliminative paraphrases.

13. For more detailed discussion see Horgan 1994c, 1995b, 1998b, 2000, and Potrc, forthcoming. In Horgan 1995b, various claims involving the semantics and metaphysics of vagueness are formulated as additional numbered theses within contextual semantics, over and above the core theses of the framework. What I say about vagueness here is largely excerpted from Horgan 1998b, 2000.

14. For convenience of exposition, I here discuss boundaryless metalinguistically, in terms of statements and their semantic status. But the same core idea applies equally well at the first-order level of description. Consider a sorites sequence consisting of the respective sand conglomerations themselves. To say that heap-hood is boundaryless is to say that (i) initially in this sequence there are heaps (with each predecessor of a heap being a heap), (ii) later there are nonheaps (with each successor of a nonheap being a nonheap), and (iii) there is no determinate fact of the matter about the transition from heaps to nonheaps.

15. The weak logical incoherence of vagueness is *generic*, in the following sense: it is not directly linked to, and does not presuppose, any particular approach to the logic of vagueness. Debates about the specific logical principles governing vague discourse are largely independent of the generic weak logical incoherence of vagueness. Strong logical incoherence is also a generic notion, not tied to any specific system of logical principles.

16. Does this mean that under the correct collective assignment of semantic status to all the statements in a sorites sequence, no two adjacent statements differ in semantic status? No. According to the collectivistic-pole requirements, there is *no* correct collective assignment of semantic status to all the statements in the sequence.

17. One reason for the name transvaluationism is to emphasize that this position is not a species of what Williamson calls *nihilism*—the view that “vague expres-

sions are empty; any vaguely drawn distinction is subverted" (1994, 165). Another reason is to emphasize the need for a "transvaluation of all truth values," so to speak, i.e., the need to transcend the impossible goal of finding some logically coherent, semantically correct, collective assignment of semantic status to all the statements in a sorites sequence. The proper goal for a semantics of vagueness, rather, is to provide an adequate account of the normative standards governing semantically correct *assertoric practice*.

18. Perhaps transvaluationism can even be implemented by standard two-valued logic, employed in a way that respects in practice the logically disciplined weak incoherence of vagueness. On accommodating vagueness, Quine remarks, "What I call my desk could be equated indifferently with countless almost coextensive aggregates of molecules, but I refer to it as a unique one of them, and I do not and cannot care which. Our standard logic takes this ... in stride, imposing a tacit fiction of unique though unspecifiable reference" (1995, 57).

19. The weak logical incoherence that any such account must take on board, at least implicitly, will inevitably reveal itself when one considers what the advocate of the particular account will be forced to say when confronted with a forced march through a sorites sequence. Consider, for instance, a sorites sequence for baldness: "A man with no hairs on his head is bald"; "A man with 1 hair on his head is bald"; ...; "A man with 10 million hairs on his head is bald." A forced march through this sequence is a series of questions, one for each successive statement, "Is it true?" Each of the questions is perfectly meaningful. And for no two successive questions could it be correct to give different answers, for that difference would mark a determinate semantic transition, contrary to the nature of vagueness. So the only thing to do, when confronted with the prospect of forced-march querying, is steadfastly to refuse to answer those persistent queries (since no complete set of answers is semantically correct). This is the *right* thing to do, because it reflects the dominance of collectivistic-pole semantic requirements over individualistic-pole requirements. But although this refusal to take the forced march is entirely appropriate as a tactic for avoiding commitment to any logically contradictory statements, it does not eliminate the weak logical incoherence of vagueness. The individualistic-pole requirements are still in force, even though they are dominated by the logically incompatible collectivistic-pole requirements, for the respective queries in the forced march are all still meaningful, and each still demands the same answer as its predecessor, even though it is proper and respectable to duck those cumulative individualistic semantic requirements by refusing to take the forced march.

20. One salient example, discussed in Horgan 1998b, section 4, is what is there called "iterated supervaluationism." The core idea of this approach is that the metalanguage for stating supervaluationist semantics is itself vague, and thus it too is subject to a supervaluationist treatment in a meta-metalanguage, and so on, all the way up the metalinguistic hierarchy. A different approach to the logic of vagueness, also a species of the transvaluationist genus, is described in Horgan 1994c, a paper that is explicit about the need to quarantine the (weak) logical incoherence that is endemic to genuine vagueness.



21. For articulation and defense of a version of indirect-correspondence metaphysical realism that countenances only one concrete particular, namely, the entire physical universe (the "BLOBJECT") construed as containing no genuine PARTS, see Horgan 1991 and Horgan and Potrc, in press.
22. Peter Unger was right! See Unger 1979a, 1979b.

## References

- Carnap, R. 1950. "Empiricism, Semantics, and Ontology." *Revue Internationale de Philosophie* 11: 20–40.
- Horgan, T. 1986a. "Psychologism, Semantics, and Ontology." *Noûs* 20: 21–31.
- Horgan, T. 1986b. "Truth and Ontology." *Philosophical Papers* 15 (1986): 1–21.
- Horgan, T. 1991. "Metaphysical Realism and Psychologicistic Semantics." *Erkenntnis* 34: 297–322.
- Horgan, T. 1994a. "Naturalism and Intentionality." *Philosophical Studies* 76: 301–326.
- Horgan, T. 1994b. "Reply to Egan." *Philosophical Studies* 76: 339–347.
- Horgan, T. 1994c. "Robust Vagueness and the Forced-March Sorites Paradox." *Philosophical Perspectives* 8: 159–188.
- Horgan, T. 1995a. Critical Study of C. Wright, *Truth and Objectivity*. *Noûs* 29: 127–138.
- Horgan, T. 1995b. "Transvaluationism: A Dionysian Approach to Vagueness." *Southern Journal of Philosophy* 33 (Spindel Conference suppl.): 97–125.
- Horgan, T. 1996. "The Perils of Epistemic Reductionism." *Philosophy and Phenomenological Research* 66: 891–897.
- Horgan, T. 1998a. "Actualism, Quantification, and Compositional Semantics." *Philosophical Perspectives* 12: 503–509.
- Horgan, T. 1998b. "The Transvaluationist Conception of Vagueness." *Monist* 81: 316–333.
- Horgan, T. 2000. "Facing Up to the Sorites Paradox." In A. Anamori ed., *Proceedings of the Twentieth World Congress of Philosophy*. Vol. 6: *Analytic Philosophy and Logic*, pp. 99–111. Philosophy Documentation Center.
- Horgan, T., and Potrc, M. 2000. "Blobjectivism and Indirect Correspondence." *Facta Philosophica* 2: 249–269.
- Horgan, T., and Timmons, M. 1993. "Metaphysical Naturalism, Semantic Normativity, and Meta-semantic Irrealism." *Philosophical Issues* 4: 180–203.
- Horgan, T., and Timmons, M. 2000. "Nondescriptivist Cognitivism: Prolegomenon to a New Metaethic." *Philosophical Papers* 29, in press.
- Lewis, D. 1979. "Scorekeeping in a Language Game." *Journal of Philosophical Logic* 8: 339–359. Reprinted in his *Philosophical Papers*, vol. 3. Oxford: Oxford University Press, 1983.

Potrc, M. Forthcoming. "Vagueness Is Robust."

Putnam, H. 1981. *Reason, Truth, and History*. Cambridge: Cambridge University Press.

Putnam, H. 1983. *Realism and Reason*. vol. 3 of *Philosophical Papers*. Cambridge: Cambridge University Press.

Quine, W. V. O. 1995. *From Stimulus to Science*. Cambridge: Harvard University Press.

Sainsbury, R. M. 1990. "Concepts without Boundaries." Inaugural lecture, King's College, London. Reprinted in R. Keefe and P. Smith, eds., *Vagueness: A Reader*. Cambridge: MIT Press, 1996.

Sellars, W. 1963. *Science, Perception, and Reality*. London: Routledge and Kegan Paul.

Sellars, W. 1968. *Science and Metaphysics*. London: Routledge and Kegan Paul.

Timmons, M. 1999. *Morality without Foundations: A Defense of Ethical Contextualism*. Oxford: Oxford University Press.

Unger, P. 1979a. "I Do Not Exist." In G. F. Macdonald, ed., *Perception and Identity*. London: Macmillan.

Unger, P. 1979b. "There Are No Ordinary Things." *Synthese* 41: 117–154.

Williamson, T. 1994. *Vagueness*. London: Routledge.

Wright, C. 1987. *Realism, Meaning, and Truth*. Oxford: Basil Blackwell.

Wright, C. 1992. *Truth and Objectivity*. Cambridge: Harvard University Press.



## II

---

### Coherence Theories



---

## Coherence Theories: Introduction

Coherence theories of truth rose to prominence at the end of the nineteenth century under the influence of the neo-Hegelian absolute idealists H. H. Joachim and F. H. Bradley. In contrast to the correspondence theory of truth, the absolute idealists denied that a belief is true because it represents the intrinsic features of reality. Truth is not a relation between a proposition and an independent realm of objects but, in Joachim's words, "the systematic coherence which characterized a significant whole" (1906, 79).

The core of absolute idealism, which received its most crisp and comprehensive presentation in the hands of the American philosopher Brand Blanshard, is a particular way of looking at the relationship of thought to reality. Rejecting the traditional distinction between subject and object, the idealist holds instead that "to think of a thing is to get that thing itself in some degree within the mind." In the idealist view, a thought and its object do not differ in kind but *in degree of realization*; the purpose of thought is to become more developed and coherent until it literally is identical to, or "one with," reality. Hence, reality *just is* the realization of a fully articulated and maximally coherent system of judgements (a "significant whole"), and a particular judgement is true just when it belongs to such a system. In short, a judgement is true when and only when it is a member of an ideally coherent system of judgments.

There are several points to note about this theory. First, by "coherence" Blanshard has in mind not simply consistency but a much richer notion. At the ideal, a coherent system is (a) one that is comprehensive, or including all "known facts," and (b) one where the support between judgements is such that "every judgement entailed, and was entailed by

the system.” Of course, our human belief systems are rarely ideal. But coherence is a matter of degree, and even if a belief isn’t absolutely true unless it is part of an ideal system, it can be partly true to the extent that is a member of a fairly coherent system.

Because the coherence theory stresses that truth consists in the coherence, or mutual support, of our beliefs, it is common to label the coherence theory as an “epistemic” conception of truth. Yet unlike some of his predecessors, Blanshard acknowledges that a coherence theory of justification (what Blanshard would call a theory of the test for truth) is logically distinct from a coherence theory of truth itself.<sup>1</sup> But while Blanshard acknowledges that coherence theories of truth and of justification are distinct, he argues that “coherence is a pertinacious concept, and, like the well-known camel, if one lets it get its nose under the edge of the tent, it will shortly walk off with the whole.”

Like Blanshard, contemporary epistemologist Linda Martín Alcoff holds coherence theories of justification and truth. On Alcoff’s version, however, to claim that a belief is true when it is a member of a coherent system is not to commit oneself to idealism or even an epistemic concept of truth. For the various elements of the relevant system are not restricted to beliefs but also include nonsubjective aspects of lived reality, such as institutional practices, cultural traditions, and the events of our everyday life. As a result, Alcoff argues, her version of the coherence theory of truth is consistent with what she calls “immanent realism,” according to which truth is neither completely subjective nor a completely objective correspondence with the world as it is in itself.

Nonetheless, Alcoff sees immanent realism as compatible with the intuition that there is more than one true story of the world. This is because systems of knowledge are at least partly the product of one’s historical situation—of one’s subject position, one’s context. Thus truth is relative in some sense. In order to avoid making truth overly subjective, however, Alcoff does not take this relativity to be a feature of the concept of truth. Rather, the relation to the historical context is built into the very content of the individual truths—or the beliefs that are true or false. Thus two seemingly contradictory beliefs  $b_1$  and  $b_2$  might remain “partially” consistent not because their truth is relative to different contexts but because their very content—what  $b_1$  and  $b_2$  are about, in other

words—is determined by context. Yet one might argue that this move avoids radical relativism only at the price of weakening the sense in which there can be more than one true account of the world. For it implies only that there can be true descriptions of different aspects of the world, not that there can be more than one true description of the same aspect of the world.

A common objection to any sort of coherence theory, first raised by Russell (chap. 1), is that it allows any proposition to be true, since any proposition can be a member of some coherent set or other. This seems anti-intuitive in the extreme, for it implies that clearly false propositions, such as the proposition that Boston is the capital of Mississippi, may be true. In his essay, Ralph Walker defends the coherence theory of truth against this and several other common objections. In particular, Walker points out that the above objection misses the point, since coherence theorists hold that truth consists in the coherence of *some specific set of beliefs that actually are or would be held*, not in the coherence of propositions considered in the abstract. Thus it is irrelevant that the proposition that Boston is the capital of Mississippi coheres with some set of propositions.

Yet Walker ultimately argues that the very feature of coherence theories that makes them immune to Russell's objection proves to be their ultimate undoing. For consider the belief *that belief b is actually held*. The truth of this belief will consist in its coherence within the system of beliefs. Thus whether belief *b* is true depends in part on whether *b* is an actual belief, and (for any belief whatsoever) whether it is an actual belief depends only on whether *the belief that b is an actual belief* is true (or coheres with the system). Thus it would seem that there is no way to ground or restrict membership in a coherent system. There is no independent way, outside of coherence, of determining which beliefs are actual beliefs, and therefore which beliefs are true. So it seems that the worry behind Russell's objection remains.

## Note

1. Coherence theories of justification hold that a belief is epistemically justified just when it is a member of a coherent system of beliefs. Two important defenses of epistemic coherentism are Lehrer 1990 and Bonjour 1985.



**Further Reading for Part II**

- Alcoff, L. 1995. *Real Knowing: New Versions of the Coherence Theory*. Ithaca: Cornell University Press.
- Alcoff, L. 1997. "Reply to Critics." *Social Epistemology* 12, no. 3: 265–285.
- Bonjour, L. 1985. *The Structure of Empirical Knowledge*. Cambridge: Harvard University Press.
- Bradley, F. H. 1897. *Appearance and Reality*. 2nd ed. Oxford: Clarendon Press.
- Bradley, F. H. 1914. *Essays on Truth and Reality*. Oxford: Clarendon Press.
- Candish, S. 1989. "The Truth about F. H. Bradley." *Mind* 98: 331–348.
- Dauer, Francis. 1974. "In Defense of the Coherence Theory of Truth." *Journal of Philosophy* 71: 791–811.
- Davidson, D. 1986. "A Coherence Theory of Truth and Knowledge." In E. LePore, ed., *Truth and Interpretation*. Oxford: Blackwell.
- Joachim, H. H. 1906. *The Nature of Truth*. Oxford: Clarendon Press.
- Lehrer, K. 1990. *Theory of Knowledge*. Boulder: Westview Press.
- Neurath, O. 1931. "Sociology and Physicalism." A translation appears in A. J. Ayer, ed., *Logical Positivism*, pp. 282–517. Glencoe: Free Press, 1959.
- Rescher, N. 1973. *The Coherence Theory of Truth*. Oxford: Oxford University Press.
- Russell, B. 1966. "The Monistic Theory of Truth." In his *Philosophical Essays*. New York: Simon and Schuster.
- Walker, R. 1989. *The Coherence Theory of Truth: Realism, Anti-realism, Idealism*. London: Routledge.
- Wright, C. 1999. "Truth: A Traditional Debate Reviewed." In Simon Blackburn and Keith Simmons, eds., *Truth*. Oxford: Oxford University Press.

---

## Coherence as the Nature of Truth

Brand Blanshard

### 1

It has been contended in the last chapter that coherence is in the end our sole criterion of truth. We have now to face the question whether it also gives us the nature of truth. We should be clear at the beginning that these are different questions, and that one may reject coherence as the definition of truth while accepting it as the test. It is conceivable that one thing should be an accurate index of another and still be extremely different from it. There have been philosophers who held that pleasure was an accurate gauge of the amount of good in experience, but that to confuse good with pleasure was a gross blunder. There have been a great many philosophers who held that for every change in consciousness there was a change in the nervous system and that the two corresponded so closely that if we knew the laws connecting them we could infallibly predict one from the other; yet it takes all the hardihood of a behaviourist to say that the two are the same. Similarly it has been held that though coherence supplies an infallible measure of truth, it would be a very grave mistake to identify it with truth.

### 2

The view that truth *is* coherence rests on a theory of the relation of thought to reality, and since this is the central problem of the theory of knowledge, to begin one's discussion by assuming the answer to it or by trying to make one out of whole cloth would be somewhat ridiculous. But as this was our main problem in the long discussions of Book II,

we may be pardoned here for brevity. First we shall state in *résumé* the relation of thought to reality that we were there driven to accept, and sketch the theory of truth implicit in it. We shall then take up one by one the objections to this theory and ask if they can pass muster.

To think is to seek understanding. And to seek understanding is an activity of mind that is marked off from all other activities by a highly distinctive aim. This aim, as we saw in our chapter on the general nature of understanding, is to achieve systematic vision, so to apprehend what is now unknown to us as to relate it, and relate it necessarily, to what we know already. We think to solve problems; and our method of solving problems is to build a bridge of intelligible relation from the continent of our knowledge to the island we wish to include in it. Sometimes this bridge is causal, as when we try to explain a disease; sometimes teleological, as when we try to fathom the move of an opponent over the chess board; sometimes geometrical, as in Euclid. But it is always systematic; thought in its very nature is the attempt to bring something unknown or imperfectly known into a sub-system of knowledge, and thus also into that larger system that forms the world of accepted beliefs. That is what explanation is. *Why* is it that thought desires this ordered vision? *Why* should such a vision give satisfaction when it comes? To these questions there is no answer, and if there were, it would be an answer only because it had succeeded in supplying the characteristic satisfaction to this unique desire.

But may it not be that what satisfies thought fails to conform to the real world? Where is the guarantee that when I have brought my ideas into the form my ideal requires, they should be *true*? Here we come round again to the tortured problem of Book II. In our long struggle with the relation of thought to reality we saw that if thought and things are conceived as related only externally, then knowledge is luck; there is no necessity whatever that what satisfies intelligence should coincide with what really is. It may do so, or it may not; on the principle that there are many misses to one bull's-eye, it more probably does not. But if we get rid of the misleading analogies through which this relation has been conceived, of copy and original, stimulus and organism, lantern and screen, and go to thought itself with the question what reference to an object means, we get a different and more hopeful answer. To think of a

thing is to get that thing itself in some degree within the mind. To think of a colour or an emotion is to have that within us which if it *were developed and completed*, would identify itself with the object. In short, if we accept its own report, thought is related to reality as the partial to the perfect fulfilment of a purpose. The more adequate its grasp the more nearly does it approximate, the more fully does it realize in itself, the nature and relations of its objects.

### 3

Thought thus appears to have two ends, one immanent, one transcendent. On the one hand it seeks fulfilment in a special kind of satisfaction, the satisfaction of systematic vision. On the other hand it seeks fulfilment in its object. Now it was the chief contention of our second book that these ends are one. Indeed unless they are accepted as one, we could see no alternative to scepticism. If the pursuit of thought's own ideal were merely an elaborate self-indulgence that brought us no nearer to reality, or if the apprehension of reality did not lie in the line of thought's interest, or still more if both of these held at once, the hope of knowledge would be vain. Of course it may really be vain. If anyone cares to doubt whether the framework of human logic has any bearing on the nature of things, he may be silenced perhaps, but he cannot be conclusively answered. One may point out to him that the doubt itself is framed in accordance with that logic, but he can reply that thus we are taking advantage of his logico-centric predicament; further, that any argument we can offer accords equally well with his hypothesis and with ours, with the view that we are merely flies caught in a logical net and the view that knowledge reveals reality. And what accords equally well with both hypotheses does not support either to the exclusion of the other. But while such doubt is beyond reach by argument, neither is there anything in its favour.<sup>1</sup> It is a mere suspicion which is, and by its nature must remain, without any positive ground; and as such it can hardly be discussed. Such suspicions aside, we can throw into the scale for our theory the impressive fact of the advance of knowledge. It has been the steadfast assumption of science whenever it came to an unsolved problem that there was a key to it to be found, that if things happened thus rather than

otherwise they did so for a cause or reason, and that if this were not forthcoming it was never because it was lacking, but always because of a passing blindness in ourselves. Reflection has assumed that pursuit of its own immanent end is not only satisfying but revealing, that so far as the immanent end is achieved we are making progress toward the transcendent end as well. Indeed, that these ends coincide is the assumption of every act of thinking whatever. To think is to raise a question; to raise a question is to seek an explanation; to seek an explanation is to assume that one may be had; so to assume is to take for granted that nature in that region is intelligible. Certainly the story of advancing knowledge unwinds as if self-realization in thought meant also a coming nearer to reality.

## 4

That these processes are really one is the metaphysical base on which our belief in coherence is founded. If one admits that the pursuit of a coherent system has actually carried us to what everyone would agree to call knowledge, why not take this ideal as a guide that will conduct us farther? What better key can one ask to the structure of the real? Our own conviction is that we should take this immanent end of thought in all seriousness as the clue to the nature of things. We admit that it may prove deceptive, that somewhere thought may end its pilgrimage in frustration and futility before some blank wall of the unintelligible. There are even those who evince their superior insight by taking this as a foregone conclusion and regarding the faith that the real is rational as the wishful thinking of the 'tender-minded'. Their attitude appears to us a compound made up of one part timidity, in the form of a refusal to hope lest they be disillusioned; one part muddled persuasion that to be sceptical is to be sophisticated; one part honest dullness in failing to estimate rightly the weight of the combined postulate and success of knowledge; one part genuine insight into the possibility of surds in nature. But whatever its motives, it is a view that goes less well with the evidence than the opposite and brighter view. That view is that reality is a system, completely ordered and fully intelligible, with which thought in its advance is more and more identifying itself. We may look at the growth of knowledge,

individual or social, either as an attempt by our own minds to return to union with things as they are in their ordered wholeness, or the affirmation through our minds of the ordered whole itself. And if we take this view, our notion of truth is marked out for us. Truth is the approximation of thought to reality. It is thought on its way home. Its measure is the distance thought has travelled, under guidance of its inner compass, toward that intelligible system which unites its ultimate object with its ultimate end. Hence at any given time the degree of truth in our experience as a whole is the degree of system it has achieved. The degree of truth of a particular proposition is to be judged in the first instance by its coherence with experience as a whole, ultimately by its coherence with that further whole, all-comprehensive and fully articulated, in which thought can come to rest.

5

But it is time we defined more explicitly what coherence means. To be sure, no fully satisfactory definition can be given; and as Dr. Ewing says, 'it is wrong to tie down the advocates of the coherence theory to a precise definition. What they are doing is to describe an ideal that has never yet been completely clarified but is none the less immanent in all our thinking.'<sup>2</sup> Certainly this ideal goes far beyond mere consistency. Fully coherent knowledge would be knowledge in which every judgement entailed, and was entailed by, the rest of the system. Probably we never find in fact a system where there is so much of interdependence. What it means may be clearer if we take a number of familiar systems and arrange them in a series tending to such coherence as a limit. At the bottom would be a junk-heap, where we could know every item but one and still be without any clue as to what that remaining item was. Above this would come a stone-pile, for here you could at least infer that what you would find next would be a stone. A machine would be higher again, since from the remaining parts one could deduce not only the general character of a missing part, but also its special form and function. This is a high degree of coherence, but it is very far short of the highest. You could remove the engine from a motor-car while leaving the other parts intact, and replace it with any one of thousands of other engines, but the thought of such an

interchange among human heads or hearts shows at once that the interdependence in a machine is far below that of the body. Do we find then in organic bodies the highest conceivable coherence? Clearly not. Though a human hand, as Aristotle said, would hardly be a hand when detached from the body, still it would be something definite enough; and we can conceive systems in which even this something would be gone. Abstract a number from the number series and it would be a mere unrecognizable  $x$ ; similarly, the very thought of a straight line involves the thought of the Euclidean space in which it falls. It is perhaps in such systems as Euclidean geometry that we get the most perfect examples of coherence that have been constructed. If any proposition were lacking, it could be supplied from the rest; if any were altered, the repercussions would be felt through the length and breadth of the system. Yet even such a system as this falls short of ideal system. Its postulates are unproved; they are independent of each other, in the sense that none of them could be derived from any other or even from all the others together; its clear necessity is bought by an abstractness so extreme as to have left out nearly everything that belongs to the character of actual things. A completely satisfactory system would have none of these defects. No proposition would be arbitrary, every proposition would be entailed by the others jointly and even singly,<sup>3</sup> no proposition would stand outside the system. The integration would be so complete that no part could be seen for what it was without seeing its relation to the whole, and the whole itself could be understood only through the contribution of every part.

## 6

It may be granted at once that in common life we are satisfied with far less than this. We accept the demonstrations of the geometer as complete, and do not think of reproaching him because he begins with postulates and leaves us at the end with a system that is a skeleton at the best. In physics, in biology, above all in the social sciences, we are satisfied with less still. We test judgements by the amount of coherence which in that particular subject-matter it seems reasonable to expect. We apply, perhaps unconsciously, the advice of Aristotle, and refrain from asking demonstration in the physical sciences, while in mathematics we refuse to

accept less. And such facts may be thought to show that we make no actual use of the ideal standard just described. But however much this standard may be relaxed within the limits of a particular science, its influence is evident in the grading of the sciences generally. It is precisely in those sciences that approach most nearly to system as here defined that we achieve the greatest certainty, and precisely in those that are most remote from such system that our doubt is greatest whether we have achieved scientific truth at all. Our immediate exactions shift with the subject-matter; our ultimate standard is unvarying.

## 7

Now if we accept coherence as the test of truth, does that commit us to any conclusions about the *nature* of truth or reality? I think it does, though more clearly about reality than about truth. It is past belief that the fidelity of our thought to reality should be rightly measured by coherence if reality itself were not coherent. To say that the nature of things may be *incoherent*, but we shall approach the truth about it precisely so far as our thoughts become coherent, sounds very much like nonsense. And providing we retained coherence as the test, it would still be nonsense even if truth were conceived as correspondence. On this supposition we should have truth when, our thought having achieved coherence, the correspondence was complete between that thought and its object. But complete correspondence between a coherent thought and an incoherent object seems meaningless. It is hard to see, then, how anyone could consistently take coherence as the test of truth unless he took it also as a character of reality.

## 8

Does acceptance of coherence as a test commit us not only to a view about the structure of reality but also to a view about the nature of truth? This is a more difficult question. As we saw at the beginning of the chapter, there have been some highly reputable philosophers who have held that the answer to 'What is the test of truth?' is 'Coherence', while the answer to 'What is the nature or meaning of truth?' is 'Correspon-



dence'. These questions are plainly distinct. Nor does there seem to be any direct path from the acceptance of coherence as the test of truth to its acceptance as the nature of truth. Nevertheless there is an indirect path. If we accept coherence as our test, we must use it everywhere. We must therefore use it to test the suggestion that truth *is* other than coherence. But if we do, we shall find that we must reject the suggestion as leading to *incoherence*. Coherence is a pertinacious concept and, like the well-known camel, if one lets it get its nose under the edge of the tent, it will shortly walk off with the whole.

Suppose that, accepting coherence as the test, one rejects it as the nature of truth in favour of some alternative; and let us assume, for example, that this alternative is correspondence. This, we have said, is incoherent; why? Because if one holds that truth is correspondence, one cannot intelligibly hold either that it is tested by coherence or that there is any dependable test at all. Consider the first point. Suppose that we construe experience into the most coherent picture possible, remembering that among the elements included will be such secondary qualities as colours, odours, and sounds. Would the mere fact that such elements as these are coherently arranged prove that anything precisely corresponding to them exists 'out there'? I cannot see that it would, even if we knew that the two arrangements had closely corresponding patterns. If on one side you have a series of elements  $a, b, c, \dots$ , and on the other a series of elements  $\alpha, \beta, \gamma, \dots$ , arranged in patterns that correspond, you have no proof as yet that the *natures* of these elements correspond. It is therefore impossible to argue from a high degree of coherence within experience to its correspondence in the same degree with anything outside. And this difficulty is typical. If you place the nature of truth in one sort of character and its test in something quite different, you are pretty certain, sooner or later, to find the two falling apart. In the end, the only test of truth that is not misleading is the special nature or character that is itself constitutive of truth.

Feeling that this is so, the adherents of correspondence sometimes insist that correspondence shall be its own test. But then the second difficulty arises. If truth does consist in correspondence, no test can be sufficient. For in order to know that experience corresponds to fact, we

must be able to get at that fact, unadulterated with idea, and compare the two sides with each other. And we have seen in the last chapter that such fact is not accessible. When we try to lay hold of it, what we find in our hands is a judgement which is obviously not itself the indubitable fact we are seeking, and which must be checked by some fact beyond it. To this process there is no end. And even if we did get at the fact directly, rather than through the veil of our ideas, that would be no less fatal to correspondence. This direct seizure of fact presumably gives us truth, but since that truth no longer consists in correspondence of idea with fact, the main theory has been abandoned. In short, if we can know fact only through the medium of our own ideas, the original forever eludes us; if we can get at the facts directly, we have knowledge whose truth is not correspondence. The theory is forced to choose between scepticism and self-contradiction.<sup>4</sup>

Thus the attempt to combine coherence as the test of truth with correspondence as the nature of truth will not pass muster by its own test. The result is *incoherence*. We believe that an application of the test to other theories of truth would lead to a like result. The argument is: assume coherence as the test, and you will be driven by the incoherence of your alternatives to the conclusion that it is also the nature of truth.

The theory that truth *consists* in coherence must now be developed more specifically. The theory has been widely attacked, and the average reader will not improbably come to it with numerous and dark suspicions. In presenting the theory we shall therefore follow a somewhat unusual procedure. We shall go down the line of these suspicions and objections, trying to deal with them in roughly the order in which they naturally arise, and seeking in our answers to bring the nature and implications of the theory gradually to light.

## 9

### (1)

It is objected, first, that the view entails scepticism. What is it that our judgements must cohere with in order to be true? It is a system of knowledge complete and all-inclusive. But obviously that is beyond us—

very probably forever beyond us. If to know anything as true, which means simply to know it, requires that we should see its relation to the total of possible knowledge, then we neither do nor can know anything.

The answer lies partly in an admission, partly in an explanation. The admission is that the theory does involve a degree of scepticism regarding our present knowledge and probably all future knowledge. In all likelihood there will never be a proposition of which we can say, 'This that I am asserting, with precisely the meaning I now attach to it, is absolutely true'. Such a conclusion may bring disappointment, but disappointment is not discredit. And in the light of the history of science, this refusal to claim absoluteness for our knowledge appears even as a merit. For the road of history is so thick with discarded certainties as to suggest that any theory which distributes absolute guarantees is touched with charlatanism. Those who would define truth as correspondence or self-evidence commonly believe that in certain judgements these characters can be found to the full and hence that the judgements are true absolutely. But it is easy to point to past judgements which, in the best opinion of the time, satisfied both definitions at once—judgements for example about the flatness of the earth or the rising of the sun—which nevertheless turned out false. In the light of such facts, theories that give patents of absoluteness to any of our present truths have antecedent probability against them. It may be answered that if judgements seeming to be true have turned out false, this does not show that truth has been wrongly defined but only that men have made a mistake as to whether its defining character was present. But the answer is obvious. The objection now before us is that, in contrast with other theories, coherence leads to scepticism. If it is now admitted that the other theories themselves are so difficult to apply that one can have no certainty, even in leading cases, whether the character they define as truth is present or not, then these theories are sceptical also.

We may reply, secondly, with an explanation, which comes essentially to this, that the coherence theory, like other theories, needs to be applied with some common sense. While the truth of a judgement does consist in the last resort in its relations to a completed system, no sensible person would claim to know these in detail, or deny the judgement *any* truth till

he did know them, any more than he would deny some beauty to a picture because it failed of beauty absolute. The system we actually work with is always less than *the* whole; at the best it is the mass of scientific knowledge bearing on the point in question; on the average it is a cloudy congeries of memories, suggestions and inferences, ill-organized in the extreme, and yet capable of subconscious mobilization and use. And for all of us, except in rare moments, the interest in truth is satisfied by exercise within these limits. Even the scientist is commonly satisfied if his theory receives the *imprimatur* of the organized knowledge of his time, and he would think it fantastic to attack him on the ground that organized knowledge has been known to change, that it may do so again, and hence that his theory may have to change with it. This last he would no doubt admit, adding however that to allow one's pursuit of science, or one's confidence in it, to be practically affected by this is merely silly. We agree. For all the ordinary purposes of life, coherence does not mean coherence with some inaccessible absolute, but with the system of present knowledge; and since this is by no means beyond determining, to describe the theory as simply sceptical is misleading. In practice it is not sceptical at all; in theory it upholds the scepticism that is a mainspring of progress. It justifies our acceptance of beliefs scientifically tested, while providing a salutary warning that science itself may become a fetish. While supporting the belief in scientific advance, it refuses to believe that this advance has reached the end of the road. It is absolutistic without dogmatism, and relativistic without countenancing despair.

## 10

### (2)

This answers by implication another objection to the theory. It is said that a truth once true must be always true, whereas on the coherence theory what *was* true may now be false, and what is now true may become false with expanding knowledge. That which coheres with the knowledge of an earlier time may conflict with the knowledge of a later time. Thus propositions may put on truth or falsity, and take them off again, with changing scientific fashions; which is absurd.

But the objection is baseless. The measure of truth, which, judged by the ultimate standard, belongs to the proposition 'x is y' is quite unalterable, for the coherence theory as for its critics. But as just admitted, we cannot in practice make use of that ultimate standard, and are compelled to fall back on a second best. What the ultimate standard means *in practice* is the system of present knowledge as apprehended by a particular mind. That system changes; hence what coheres with it at one time may not cohere with it at another; thus in practice we shall be justified in accepting at one time what later we must reject. This is all true, but where is the inconsistency? We have neither said nor implied that truth itself changes. What we have said is that while truth as measured by the ultimate standard is unchanging, our knowledge of that truth does change—which is a very different thing. Our system of knowledge fluctuates; it is not now, for example, what it was in the Dark Ages, or even in the middle of the last century; and if we use as our standard this variable measuring-rod we shall naturally get varying results. But these varying results are in our knowledge, or in truth-as-revealed-in-our-knowledge, not in truth objective and complete. Between a truth that is itself invariant and varying degrees of manifestation of this truth, there is no sort of inconsistency.

## 11

### (3)

This answer suggests a third objection. We have held that while the truth of any particular proposition must be tested by its coherence with present knowledge, the truth of this knowledge as a whole could be measured only by its approximation to an absolute system. But it has been charged that 'approximation' covers a surrender to correspondence.<sup>5</sup> For do we not really mean by this that our present system is true so far as it *corresponds* to the further reality, and false so far as it fails of this?

We may call the relation 'correspondence' if we wish. Indeed some of the most uncompromising advocates of coherence have used the language of correspondence in their discussions of this point; Bradley, for example, speaks of our judgements as 'representatives' of reality which

are true 'just so far as they agree with, and do not diverge from', the real.<sup>6</sup> Again, 'truth, to be true, must be true of something, and this something itself is not truth. This obvious view I endorse.'<sup>7</sup> But he adds, 'to ascertain its proper meaning is not easy'. And what he arrives at as the 'proper meaning' is certainly very far from correspondence as meant by its advocates. It is neither copying, nor a one-to-one relation, nor an indefinable 'accordance'; 'I mean', he writes of judgements, 'that less or more they actually possess the character and type of absolute truth and reality. They can take the place of the Real to varying extents, because containing in themselves less or more of its nature. They are its representatives, worse or better, in proportion as they present us with truth affected by greater or less derangement.' 'We may put it otherwise by saying that truths are true, according as it would take less or more to convert them into reality.'<sup>8</sup> Or, if we may put in our own terms a meaning that is certainly not far from Bradley's, the relation is one between a purpose partially fulfilled and a purpose fulfilled completely. Thought, we have insisted, *is* its object realized imperfectly, and a system of thought is true just so far as it succeeds in embodying that end which thought in its very essence is seeking to embody. If we want analogies for the relation of our thought to the system that forms its end, we should leave aside such things as mirrors and number systems and their ways of conforming to objects, and think of the relation between seed and flower, or between the sapling and the tree. Does the sapling *correspond* to the tree that emerges from it? If you say it does, we shall agree that a system of thought may correspond to reality. If, as seems far more likely, you say it does not, and that to use 'correspondence' of such a relation is confusing, then you are at one with us in considering 'correspondence' a misdescription of the relation we have in mind.

## 12

### (4)

Just as certain critics have attempted to reduce coherence to correspondence, certain others have attempted to reduce it to self-evidence. They say: 'When we grasp the coherence of a proposition with a system, we

are seeing that it necessitates and is necessitated by the other elements in the whole; and what we mean by necessary relations is relations logically self-evident.'

Again we must answer by defining terms. When anyone says he believes in self-evidence, he is commonly taken to mean that he believes in self-evident *propositions*, that is, in propositions whose truth can be seen without considering how they are related to the systems they belong to. Thus Descartes believed in self-evidence because he believed that there were certain 'simple propositions' which, however fertile of consequences when the mind reflected on them, could be seen to be true by themselves before any such consequences were deduced. This is a useful way of conceiving self-evidence, and as it is also the commonest way, it seems wisest to conform to it. But if we do, it is plain at once that to reduce coherence to self-evidence is out of the question, since the two theories contradict each other on an essential point. The self-evidence theory says the truth of some propositions at least can be seen in isolation; the coherence theory says that the truth of *no* proposition can be seen in isolation.

However, the defender of self-evidence may reject the proposed definition; he may insist that what he means by self-evidence is something attaching equally to propositions in isolation and to the coherence of these with a system. This is a distinct view and demands a distinct answer. That answer is not difficult, and it is to our mind decisive against any form of self-evidence that may be offered as an account of truth. Self-evidence, in its essence, contains a reference to being seen; if a truth were too complicated and difficult for any human apprehension, no one would call it self-evident. And if not self-evident, then on the theory it could not be a truth at all. Now this is a violent paradox. It involves the conclusion that if the best human brains cannot *see* a proposition to be true, then it cannot *be* true. It suggests that when Newton, having hit on the law of gravitation, laid this aside for a while because his calculations failed to confirm it, the law was really not true, since it possessed self-evidence for no one. It is surely more natural to believe that there are numberless truths too recondite and elaborately conditioned for human wit. So long as self-evidence is offered merely as a criterion of truth, there is some plausibility in it, as we have seen; but when offered as the nature of truth, the plausibility vanishes.

## 13

## (5)

We come now to an objection more frequently made than any we have been considering. Granting that propositions, to be true, must be coherent with each other, may they not be coherent without being true? Are there not many systems of high unity and inclusiveness, which nevertheless are false? We have seen, for example, that there are various systems of geometry each of which seems to be as coherent internally as the others. Since they are mutually inconsistent, not more than one of them can be true, and there are many mathematicians who would say that *none* of them are true; yet if truth lies merely in coherence, are we not compelled to take all of them as true? Again, a novel, or a succession of novels such as Galsworthy's *Forsyte Saga*, may create a special world of characters and events which is at once extremely complex and internally consistent; does that make it the less fictitious? To say that it does would imply that if we could only dream constantly enough and consistently enough our dreams would literally come true.

## (i)

This objection, like so many other annihilating criticisms, would have more point if anyone had ever held the theory it demolishes. But if intended to represent the coherence theory as responsibly advocated, it is a gross misunderstanding. That theory does not hold that any and every system is true, no matter how abstract and limited; it holds that one system only is true, namely the system in which everything real and possible is coherently included. How one can find in this the notion that a system would still give truth if, like some arbitrary geometry, it disregarded experience completely, it is not easy to see.

## 14

## (ii)

The objection gains point, however, when it goes on to inquire whether all that is actual might not be embraced in more than one system. When a murder is committed, there may be two theories of the crime which do



complete and equal justice to all the known facts and yet are inconsistent with each other. Is it not conceivable similarly that there should be two perfect but conflicting systems in which all known and knowable facts should fall into place? If so, our standard would require us to say that both were true; yet since they conflict, this would be absurd. Now we might reply that such a contingency, though possible, is highly improbable. In the case of the murder, every new bit of evidence narrows the range of available hypotheses, and it does not even occur to us that if we knew *all* the relevant facts we might find ourselves at the end with conflicting theories. If such an issue is improbable where the facts are so few, is it not far more improbable where the facts are infinitely many?

Still, this answer seems inadequate, since a theory that leaves it even possible that in the ultimate nature of truth there should be inconsistency ought to be met, we feel, with some decisive disproof. Can it be shown that such an issue is not only improbable, but impossible? I think it can. There are to be two systems, each including all facts known or knowable, but differing in internal structure. Now if the first system is constructed according to plan A, and the second according to plan B, then the possession by the first of plan A is not a fact that is included in the second, and the possession of plan B by the second is not a fact included in the first. The two systems are thus *not*, as they are supposed to be, each inclusive of all the known facts. To put it otherwise, if the systems differ neither in facts nor in structure, they are not two systems but one. If, with the same facts, they are to differ at all, they must differ in structure, but then there will be at least one fact that each of them must omit, namely, the fact that the other possesses the particular structure it does. Thus that all actual and possible facts should be embraced in conflicting systems is unthinkable.

On the other hand, if the objector lowers his claim and says only that the facts *as so far known* may be ordered in different systems, he is saying nothing against our theory. For this certainly does not show that if all the facts were known these rivals would still stand as rivals; it shows only that with the facts now available we should not on our view be justified in making a choice. And this really confirms our view, through bringing it into line with science. Such suspension of judgement is precisely what is enjoined by scientific practice, which holds that so long as two rival

hypotheses equally cover the facts, neither is to be preferred to the other, but that as soon as there appears an *instantia crucis* which one hypothesis can assimilate and the other not, we are justified in adopting the first.<sup>9</sup>

15

(iii)

Suppose, however, that no crucial instance ever did arise. Suppose (to put an extreme but conceivable case) that we spent from twelve midnight to twelve noon of every day in dreaming, that our dreams were as vivid and orderly as our waking life, and that when we resumed them every night we did so at exactly the point at which we left off the day before. Would there then be any difference between sleep and waking? Would there be any sense in saying that one world was real and the other unreal, that in the one our perceptions and beliefs were true and in the other delusions merely? I think not. And our inability to make any choice in such a conjuncture confirms our theory. The argument runs: if truth did lie in coherence, then, confronted with two worlds equally coherent, we should be unable to select one as truer than the other; on reflection we can see that such inability is just what we should find; hence the equation of truth with coherence is so far verified.

16

(iv)

It is further verified by our way of choosing between systems which in the above sense are *not* equal. There are various cases. Consider (a) how we recognize dreams or delusions for what they are. When we are suddenly roused from a vivid dream, we may be momentarily dazed, not knowing the dream from the actuality. How do we establish which is which? Mere vividness does not decide the matter; the dream may be of nightmare intensity while the perception of our familiar surroundings may be comparatively dim. The deciding factor in the battle is what may be called the mass and integration of the household troops. The bureau and windows of our familiar bedroom and the sound of a familiar voice throw out innumerable lines of connection that bring our everyday world around us

again in irresistible volume. Against the great bulk of this world, and without any lodgement in it, the figures of our dream appear unsubstantial and fugitive, quickly dissolving for want of support; and it is just the recognition that what we have been experiencing will not fit into our common-sense world that we mean when we say we wake from dream. The power to measure such fancies and phantasms against the ordered mass of experience is the logical meaning of sanity; its disappearance is insanity. There may be organic differences between the man who thinks himself Napoleon, the man who is sure he has committed the unpardonable sin, and the man who is persuaded that there is a universal conspiracy to keep him down; but intellectually they are alike; there are certain beliefs which resist appraisal by the mass of their general experience, and stand in the midst of it like solid capsules impervious to outer influences. In these cases that is what insanity means.<sup>10</sup>

#### Notes

1. See further, Chap. xxx, Sec. 15.
2. *Idealism*, 231.
3. Coherence can be defined without this point, which, as Dr. Ewing remarks (*Idealism*, 231), makes the case harder to establish. In no mathematical system, for example, would anyone dream of trying to deduce all the other propositions from any proposition taken singly. But when we are describing an ideal, such a fact is not decisive, and I follow Joachim in holding that in a perfectly coherent system every proposition would entail all others, if only for the reason that its meaning could never be fully understood without apprehension of the system in its entirety.
4. Cf. the criticism of the copy theory above, Chap. vii, Sec. 9. And see the appendix to the present chapter for comment on a current defence of correspondence.
5. As, for example, by Schiller, *Studies in Humanism*, 122.
6. *Appearance*, 362–3.
7. *Essays on Truth and Reality*, 325; and cf. 'If my idea is to work it must correspond to a determinate being it cannot be said to make'.
8. *Appearance*, 362–3.
9. It may be said that the truth is not established until *all* rivals have been eliminated. But this is not the view on which the natural sciences actually proceed. Of course in formal logic an argument from the affirmation of the consequent is fallacious, and when this is carried over into science it is often said to provide verification without proof; the proof is attained only when it is shown that from no

other antecedent could these consequences have sprung. But it will be evident that in the ordinary work of science proof of this kind is seldom if ever practicable; one cannot be sure that *all possible* alternatives have been excluded. 'The character of relativity and non-finality, which attaches to mere verification and causes it to be called the fallacy of the consequent, is really inevitable in the pursuit of truth.'—Bosanquet, *Implication and Lin. Inf.*, 102.

10. Much evidence could be adduced for the above suggestions as to the nature of sanity and of aberrations from it. See, e.g., McDougall's account of relative dissociation as explaining the lack of normal inhibition in hypnosis. *Abnormal Psychology*, 110 ff.



---

## The Coherence Theory

Ralph C. S. Walker

### I

The coherence theory of truth deserves better treatment than it has usually been given in the past fifty years or so. The stock dismissals of it—by Russell, by Schlick, and by many others who ought to have known better<sup>1</sup>—dispose only of a crude caricature. But they have been widely accepted as exhibiting the theory as a foolish aberration, suited only to the occasional muddled verificationist and to the Idealists of the last century (a time when, as we like to think, philosophical standards were so much lower that the most palpable absurdities escaped attention).

That these dismissals were unduly hasty is suggested by the theory's recent revival in the hands of such hard-headed philosophers as Quine, Davidson and Putnam; on a plausible interpretation of Wittgenstein it can be ascribed also to him. There are, of course, substantial differences between these contemporary philosophers and the coherence theorists of the last century, and also among these contemporary philosophers themselves. The differences are great enough to suggest that it might be better to speak of coherence theories, in the plural, rather than of 'the coherence theory' of truth. Nevertheless I think these various views have enough in common to justify the traditional usage. The motivations for them—motivations which are powerful and hard to resist—are very similar; they run into parallel difficulties; and above all they share the same general character, the same radical conception of the kind of thing that truth is. I shall therefore continue to speak (when it is not misleading to do so) of 'the coherence theory of truth' in the singular.

The first thing to be clear about is what this coherence theory is. It is a theory about the nature of truth. When Rescher wrote his book *The Coherence Theory of Truth* it seemed to him (because he accepted the stock objections) that no one could seriously have meant it in that way, and so he concerns himself instead with the view that coherence provides the *criterion* for truth—the way of finding out what is true.<sup>2</sup> That is itself an interesting view, but I shall not be considering it here, except incidentally. One can consistently hold (as Rescher himself does) that coherence provides the criterion of truth, but that the nature of truth consists in something different, a correspondence of some kind. One can also indeed combine the thesis that the nature of truth is coherence with the claim that we need some different way of finding out what is true; that is a consistent combination, but it is not so attractive a position, for reasons which will appear shortly.

The coherence theorist holds that for a proposition<sup>3</sup> to be true is for it to cohere with a certain system of beliefs. It is not just that it is true if and only if it coheres with that system; it is that the coherence, and nothing else, is what its truth consists in. In particular, truth does not consist in the holding of some correspondence between the proposition and some reality which obtains independent of anything that may be believed about it.

This is a radical thesis. It conflicts with what most of us naturally think. But it is important to notice that the coherence theorist does not depart so far from common sense as to have to deny such truisms as ‘true propositions correspond with the facts’. It is common to treat the coherence and correspondence theories of truth as though they were rivals, and so they are, if the correspondence theory is also a theory about the nature of truth: a theory to the effect that truth *does* consist in some sort of correspondence between a proposition on the one hand, and on the other a real world whose nature and existence are quite independent of what may be believed about it. But although that is a very natural theory, one does not commit oneself to it just by saying things like ‘true propositions correspond with the facts’. Coherence theorists can make such remarks quite freely; they just will not regard them as expressing the nature of truth, nor will they take ‘the facts’ to belong to a metaphysically independent reality; on the contrary, on their view the facts are them-

selves determined by the coherent system of beliefs. But ‘corresponds with the facts’ is so standardly used as a long-winded equivalent of ‘is true’ that it would seem highly perverse to deny that in some sense true propositions do correspond with the facts. What is open to controversy is how informative a remark it is, and this will much depend on how we construe the correspondence relation, and what status we accord to ‘the facts’. I shall return to this, and to how the coherence theory relates to the traditional discussions about truth and correspondence, in the next chapter. In what follows, unless I indicate otherwise I shall mean by ‘the correspondence theory’ the theory that truth consists in some kind of correspondence with a reality independent of what may be believed about it—the view that the coherence theorist is out to reject. (It would not do to characterize it as the theory that truth consists in correspondence with a reality independent of what may be believed, and leave it at that, because there are truths about beliefs, and truths about beliefs can hardly be independent of the beliefs that make them true. They may however be independent of anything that may be believed about those beliefs themselves, and that is what the correspondence theory holds. One’s beliefs about what beliefs are held may very often be correct, especially if it is one’s own beliefs that are in question, but there is an important distinction between believing that *p* and believing that it is believed that *p*.)

Before we can go any further, though, something more positive needs to be said about what the coherence theory of truth is. To start with, it is important to emphasize that what is to be cohered with is a set of *beliefs*, and some specific set of beliefs at that. One of the standard objections to the coherence theory, made for instance by Russell,<sup>4</sup> is that (on any plausible understanding of ‘coherence’) virtually any proposition can be fitted into some coherent set. The proposition ‘Bishop Stubbs was hanged for murder’ is in fact false, but one can imagine a world no less coherent than our own in which it is true; there is thus a coherent set of propositions to which it belongs, including perhaps such propositions as ‘All bishops are murderers’ and ‘Bishops are generally hanged’, just as there is a coherent set of propositions to which (the truth) ‘Bishop Stubbs died in his bed’ belongs. But the objection misses the point, because it is not being suggested that truth consists in cohering with any arbitrary set of



propositions. No coherence theorist would ever be tempted to think that the coherence relation held simply amongst propositions in the abstract, regardless of whether anyone believed them or ever would believe them, for propositions in the abstract are hardly to be distinguished from the facts that the correspondence theorist invokes. Instead coherence theorists maintain that truth consists in coherence with a set of beliefs, and some specific set of beliefs at that—though as we shall see in a moment they do not necessarily consider that these must be beliefs that are held either by ourselves or by anyone alive today. This answer to Russell's objection raises further problems, which we shall have to return to in due course; but at least the coherence theory is not eliminated straight away by so simple a criticism.

What, though, is understood by 'coherence', and what system of beliefs is intended? Since coherence theorists are a varied lot, very different answers to these questions have been offered. Often the system of beliefs has been envisaged as being, broadly speaking, our own, though since most people's beliefs are inconsistent it cannot include all the things we believe. What many coherence theorists have in mind, therefore, is some subset of our beliefs; perhaps the largest subset that is in accord internally, perhaps some subset that is particularly fundamental, or indispensable, to our thinking (including no doubt the laws of logic and the principles of inference that we use in testing hypotheses and constructing theories). Others, conscious of the radical changes and developments that have taken place in human thought, have considered it wrong to tie truth to anything that we believe at present, and have taken it to be coherence with the system of beliefs that human beings will hold at the ultimate stage of their historical development. Others, more distrustful still of human fallibility, have regarded truth as coherence with the system of beliefs held by God or the Absolute; a view which must be carefully distinguished from that of a traditional theologian who is no coherence theorist but still maintains that God knows all the truths, and has no false beliefs. For the traditional theologian God's beliefs, unlike ours, invariably correspond with the relevant facts; these facts are independent of God's beliefs about them (though they may not be independent of His creative will). For the coherence theorist on the other hand truth consists in coherence,

and if the coherence is coherence with God's beliefs then it is His beliefs that determine what the facts are.

Opinions have differed too about what is to be meant by 'coherence'. Sometimes it has been taken to be simply consistency with the basic principles that characterize the system of beliefs. Sometimes, at the other extreme, it has been held to require mutual entailment by all the propositions in question:  $p$  will cohere with  $q$  and  $r$  only if  $p$ ,  $q$  and  $r$  all entail one another. Sometimes, again, it has been left thoroughly vague what coherence is supposed to amount to. Actually we cannot really decide what is to constitute coherence until we decide which system of beliefs is appropriate, for the two questions go together. Generally the system will itself determine what coherence with it amounts to. If for example it includes the laws of logic, a set of principles of scientific inference sufficient to determine a single theory as correct when enough evidence is in, and a further set of principles adequate to determine what counts as admissible evidence, then there is no further problem as to what constitutes coherence (at least within the domain of science): propositions describing admissible evidence will cohere, as will those stating theories determined as correct by the principles of inference. A view of very much this kind is quite often held by contemporary coherence theorists, particularly those influenced by Quine. Another commonly held view makes coherence simply a matter of agreement with the considered and long-term judgment of the community—a view often, and plausibly, ascribed to Wittgenstein. It should certainly *not* be taken for granted that any coherence theory worthy of the name must necessarily incorporate the standards of coherence that at first sight seem most natural to us, or even that it must incorporate those laws that we commonly regard as the laws of logic; Hegel's, for instance, works on a very different basis. There is no reason to place any limitation on what can count as coherence: we can consider someone to hold a coherence theory of truth if they hold that truth consists in some relationship within a set of beliefs, whatever that relationship may be. Of course, coherence theorists with particularly absurd ideas as to what it might be can be particularly easily dismissed. But most real-life coherence theorists are not in that category, even if at first sight they may appear to be.

Thus what makes something a coherence theory of truth (in the broad sense in which I am using the expression) is not that it is built round some specific concept of coherence. What makes it a coherence theory is that it is a theory about the nature of truth, to the effect that for a proposition to be true is for it to fit in with some designated set of beliefs; but which set of beliefs is designated will vary from one version of the theory to another, as will the kind of fit required. There need be no suggestion in this that every truth is the content of a belief that either is or ever will be actually held, but if coherence theorists intend their account to cover truths of all kinds they will claim that every truth is the content of a belief that *would* be held if the system of beliefs were fully worked out so as to include all those that cohere. By contrast I have described the correspondence theory of truth as claiming that truth consists in correspondence with a reality that is independent of anything that may be believed about it. Putting it in that fashion brings out that the correspondence and coherence theories are exclusive alternatives, provided that they are both intended to apply to *all* truths. But they are not quite exhaustive. For one thing there is the possibility of rejecting altogether any theory of the nature of truth, on the ground that general questions about what truth consists in are too broad to be intelligible. We shall return to this idea; it is less worth taking seriously than it may seem. And there is also the possibility of a theory which combines coherence and correspondence by giving a coherence account of some kinds of truth and a correspondence account of others. On such a mixed view—which I shall call an *impure* coherence theory—the nature of reality might be determined *partly* by the system of coherent beliefs and *partly* by something else independent of it. At first sight the idea of an impure coherence theory may seem rather peculiar, but we shall find in due course that certain theories which have historically been described as coherence theories of truth are in fact impure coherence theories. An example might be a theory which offered a correspondence account of the truth of statements about our experiences, but a coherence account of the truth of the more theoretical statements which we construct on the basis of them. Another example might be a theory which gave a correspondence account of straightforward ‘factual’ truths about the world around us, but a coherence account of evaluative truths, or of truths about possibilities and necessities.<sup>5</sup>

## II

Since the coherence theory is so often dismissed out of hand, we ought to consider the reasons which are standardly adduced for rejecting it. There are five of these. At least as they stand they are all worthless. This is not to say that it may not be possible to develop more serious difficulties for the theory out of them, and that is a matter to which we shall have to return. But it is a serious mistake to think they provide grounds for the instant dismissal the theory has so commonly received. Besides these five I shall consider two further objections of a rather more subtle kind; these are equally ineffective against the theory, but by considering them we shall be able to clarify one or two remaining points about what exactly the theory claims.

One of the five standard objections we have met already. It is the objection that the coherence theory is unable to distinguish between truth and falsity because virtually every proposition (including 'Bishop Stubbs was hanged for murder') belongs to some coherent set. As we saw, this misses the point. Coherence theorists do not say that membership of any arbitrary coherent set of propositions is sufficient for truth. What they say is that truth is coherence with a certain particular set of beliefs.

The second standard objection is that the theory cannot take account of experience. Coherence theorists are not likely to be much impressed with this. They consider that they can, and do, take account of experience, and indeed they often spend a lot of time emphasizing the importance of experience in the construction of our knowledge of the world. What they reject is only the idea that taking account of it means bringing it about that our experiential beliefs match something that is given or presented to us in such a fashion that its nature is independent of beliefs about it, and the rejection of *that* picture is something they share with many philosophers who do not subscribe to the coherence theory of truth. They share it with the proponents of the coherence theory of knowledge; they share it also with those who hold that although experiential beliefs provide a foundation for our knowledge, the structure of our beliefs in general determines the character of experience. I dealt with this elsewhere.<sup>6</sup>

According to the coherence theory of truth, we have a great many beliefs about the content of our experience, but as with any other beliefs

their truth can only consist in their coherence within the system; and if that means that the system must be a rich and elaborate system, well and good. This is not incompatible with the ordinary view that their truth depends on their being caused in us in the right sort of way, by the operation of things outside us upon our sense-organs; coherence theorists have no more difficulty in accepting that than anyone else does, for they will take the appropriate causal truths to belong to the coherent system. Admittedly, *whatever* theory one holds about truth there are going to be philosophical problems to be dealt with over the way experience bears upon our knowledge; but there is no special difficulty for the coherence theory here, and no simple refutation of it is to be achieved in this fashion.

The third of the standard objections is that the theory cannot accommodate its own truth, because it must claim more for itself than just that it coheres; it must claim to be actually *true*. But this seems quite unfair. Coherence theorists hold that truth consists in coherence, and are quite prepared to say that the truth of their own theory consists in its coherence with the relevant set of beliefs. They mean nothing more than that by calling it true, but that is quite enough, for coherence is what truth is.

Someone might seek to take this objection a stage further by asking what would happen if the coherence theory did not itself belong to the coherent system of beliefs, and the correspondence theory of the nature of truth did so instead. But this again creates no serious difficulty for the coherence theorist. If it were the correspondence theory, and not the coherence theory, that cohered with the system, it would not follow that the correspondence theory was true, but it would follow that truth cannot consist in coherence; for the coherence theory cannot be true unless it is true by its own standard. That, however, is not the position, according to coherence theorists, for the coherence theory *does* cohere with the system. And there is no reason for them to worry about what *would* have been the case if the system had been different in this respect, any more than there is reason for them to worry about what the position would have been if the system had been, for example, internally inconsistent. It is not.

A similar misunderstanding gives rise to the fourth objection, to the effect that the coherence theory makes truth relative. Of course in one

sense this is perfectly true, for it does make truth depend upon the coherent system of beliefs. In exactly the same way the correspondence theory could be said to make truth 'relative to' the independently existing facts. But in neither case is there any suggestion that some alternative standard would yield some alternative and equally good kind of truth. The coherence theorist might, it is true, admit that other people or other societies might operate with a different system of beliefs which they thought of as determining the truth; the coherence theorist might go so far as to hold that it was impossible in principle for us to convince them of the error of their ways. Nevertheless their ways are in error, and it is as important to the coherence theorist as to the correspondence theorist that this is so. Truth consists in coherence not just with any system of beliefs but with a certain specific one, and anything else, any alternative system however neat and self-contained, simply is not truth.

The fifth objection that is often made, and perhaps even more frequently felt than expressed, is that the theory does so much violence to our ordinary ways of thinking and talking as to be simply absurd—or perhaps even nonsensical, because it seeks to give our ordinary words a meaning they cannot carry. Truth is not a function of what is believed but a matter of matching external reality. The existence of trees and of dinosaurs is quite independent of anyone's beliefs; there were dinosaurs around even when there was nobody to have thoughts about them, and the fact that there are trees nowadays is not dependent on there being people.

Like the others this objection fails, but some care is necessary in order to be quite clear why it fails. It misses the point because it caricatures what coherence theorists are saying. To a very large extent they are happy to speak with the vulgar. As we have seen they are quite prepared to say that true propositions are those that correspond with the facts; though they hold that it is the coherent system of beliefs that determines what the facts are. In the same way they can talk of the world as objectively real, and as independent of my beliefs about it. Clearly (since I can be mistaken) the set of beliefs that determines the truth is not to be identified with the set of beliefs that I have at present, and to recognize that is to recognize a contrast which the coherence theorist will describe as the contrast between how things seem to me and how they really, or

objectively, are. What he does maintain is that the nature of this objective reality is itself determined by the coherent system of beliefs. That certainly conflicts with our common-sense view of things, but not in a way that renders it evidently absurd.

The coherence theorist has no difficulty either over the existence of dinosaurs before there were thoughts about them. That there were dinosaurs before there were people is a familiar scientific fact, and like all scientific facts it is true (on his view) because it coheres with the system. Muddle can be caused here, as so often, by letting metaphor run away with us: the coherence theorist is sometimes thought of as saying that the facts, and hence the world and its dinosaurs, are *created* by the system of beliefs, and the conclusion is drawn that they could not have existed before any beliefs did. But this is confused. The coherent system of beliefs determines as true all sorts of statements about the remote past, including the statement that there were dinosaurs before there were people; and being determined as true, they are true, for such is the nature of truth. It has sometimes been claimed that at least the coherence theory must be in difficulty over all sorts of *specific* truths about dinosaurs which we shall never be able to verify and which no one will ever believe: there must be some number *n*, for example, for which it is true that the total number of dinosaurs that ever existed was *n*.<sup>7</sup> But this objection too is misplaced, at any rate against many forms of the theory, for the coherence theorist is under no obligation to regard as true only what we shall someday be able to verify (although admittedly *some* coherence theorists may take this line). Setting aside the possibility of saying that the coherent system of beliefs is the system of *God's* beliefs (a possibility which, as we shall see, does have its disadvantages), it is open to the coherence theorist to hold that the truth is what one *would* get to *if* one made the best possible use of certain principles of reasoning and scientific inference; and these might suffice to determine the number of dinosaurs, even if none of us will ever determine it.

Furthermore, and for similar reasons, the coherence theorist can accept the truth of counterfactuals like 'If there had been no people there would still have been trees'—a point that has been well made by Blackburn.<sup>8</sup> In the ordinary way of taking it, the counterfactual is supported by a fact of everyday science, to the effect that the existence of trees does not depend

on that of people. It will therefore come out true for the coherence theorist just as much as for anyone else, because it coheres with the body of beliefs that determines the truth. That body of beliefs is a set of propositions  $p, q, r, \dots$ ; the truth is whatever coheres with  $p, q, r, \dots$ ; and whatever  $p, q, r, \dots$  may be in detail, we can expect the generalizations of ordinary science to cohere with them, and if they do so must also the counterfactual conditionals they support.

What can cause confusion here is the thought that  $p, q, r, \dots$  determine the truth only because they are believed; and beliefs require people to have them (or at any rate believers of some kind). That consideration is sufficient, given the coherence theory, to establish that there *are* people (or at any rate believers of some kind). That there are, no one of course disputes; what is being asked is what things would have been like if there had not been. Now counterfactual conditionals are tricky, because the assessment of them depends on just what features of the actual world are assumed to be held constant in the circumstance envisaged—not all features of the actual world can be, since the conditional's antecedent is in fact false. Thus we feel a sort of puzzlement if someone produces out of the blue the statement 'If Bizet and Verdi had been compatriots Bizet would have been Italian'; we do not know how to assess it because nothing in the context or in the antecedent tells us whether Bizet's nationality should be held constant (in which case the conditional is false) or Verdi's (in which case it is true). In the *normal* way of understanding the conditional 'If there had been no people there would still have been trees', it is clear enough that the supposition is being made within the context of our ordinary theory of the world; the proper way to assess the claim is therefore to ask what that theory, or so much of it as we can combine with the truth of the antecedent, tells us, and it is easy to see (as we did above) that the conditional thus comes out true. In the highly *abnormal* context of a discussion of the coherence theory of truth, however, things can look differently. For here it is possible to view the supposition, *not* as taking for granted our ordinary theory of the world, but as putting it in question. This gives us an alternative (though except in this special context highly unnatural) way of taking the conditional. On this alternative reading, what is assumed to be held constant is the coherence theory of truth itself: the theory that truth consists in coherence with a certain set



of beliefs. This licenses the conclusion that if there were no beliefs there would be no truth at all (and thus in particular that it could not be true that there were still trees). Assuming that if there were no people there could be no beliefs, we get the result that on this interpretation 'If there were no people there would still have been trees' is *false*.

Blackburn has rightly put a great deal of emphasis on the coherence theorist's ability to admit as true statements that at first seem to express the conception of truth and reality which he in fact rejects. Calling the correspondence view that he rejects 'realism', Blackburn is able to show that coherence theorists can and should adopt a position of 'quasi-realism', which allows them 'to mimic the intellectual practices supposedly definitive of realism' without weakening in their repudiation of realism itself.<sup>9</sup> The question arises of how far this can go. Blackburn sometimes seems to be suggesting that there is *no difference* between the realist and the coherence theorist in their assignment of truth-values to statements, but this must be too strong, or it would be impossible to express the difference between the two positions. At other times he seems to be suggesting, more plausibly, that there is no difference between them in their assignment of truth-values to the things that ordinary people, as opposed to philosophers, are likely to say. This is certainly not far from the truth. But it would be wrong to suggest (as Blackburn sometimes appears to and others clearly would) that the coherence theory itself cannot be stated by using words with their ordinary *senses* and in a perfectly ordinary fashion. For it can.

It is easy to slip into the mistake of thinking that when the correspondence theory affirms, and the coherence theory denies, that truth consists in matching a reality that is independent of what may be believed about it, they are using 'independent' in a special way—giving it a peculiarly philosophical or 'transcendental' usage which must be distinguished from its everyday or 'empirical' usage. If this were the case, there would be room for serious doubt whether any clear sense had been assigned or could be assigned to the word in its alleged transcendental usage; it would be utterly unclear which features of the ordinary sense were being retained and which abandoned. The idea that this is what is going on encourages the thought that the coherence theory may be strictly unintelligible. But fortunately that is not a difficulty we have to face. What is

meant by 'independent' is just what we all normally mean by it. It may be natural to take it for granted that the facts are independent (in the ordinary sense) of any beliefs about them, but it is exactly this that coherence theorists are denying—though they are not, of course, obliged to deny that the facts are independent of *my* beliefs about them. To suppose that it is the coherent system of beliefs that determines what the facts are may be unusual, but it is not absurd or self-evidently wrong. To suppose that the facts are determined by what I happen to believe at the moment, or by the current opinions of the British people, *would* be self-evidently wrong, but no sane coherence theorist is likely to think any such thing. On the contrary, he will agree that we hold a great many beliefs that are false. What makes them false is that they fail to cohere with that set of beliefs that determines the truth.

It should be admitted that the coherence theorist's view of truth is unusual, and contrary to common sense, though not therefore absurd. It thus cannot be quite true to say that the ordinary person and the coherence theorist will not differ in their assignments of truth-values, for they will differ over the statement of the coherence theory. At the same time, the ordinary person is not likely to have occasion to formulate this statement or its denial, so that Blackburn may be right to claim that there is no difference between the realist and the coherence theorist in their assignment of truth-values to the things ordinary people are likely to say.

An objection to this claim may be raised from a different quarter. For surely, it may be said, the realist—i.e. the correspondence theorist—will accept the principle of bivalence, which says that every statement is either true or false, whereas his opponent will not. The coherence theorist will reject it because he cannot rule out the possibility that there are statements such that neither they nor their negations cohere with the system. But that is not quite right: things are not so simple.

To start with it is a mistake to suppose that realists are necessarily committed to bivalence. What they are committed to is that the nature of reality is independent of any beliefs about it, but that is compatible with the possibility that reality itself may admit of three or more truth-values. Hence there may be statements which are neither true nor false, just because of the nature of things. Dummett, who recognizes this point, claims that although realists are not necessarily committed to the view

that every statement is either true or false, they *are* committed to the view that 'any sentence on which a fully specific sense has been conferred has a determinate truth-value independently of our actual capacity to decide what that truth-value is';<sup>10</sup> but this again is wrong. It is correct to say that on the realist view a statement's truth-value is independent of our capacity to find out about it, our beliefs about it, and indeed anyone's beliefs about it. But a realist might hold that the world was intrinsically vague in such a way as to make some quite specific claims (as for example the claim that so-and-so is not yet an adult) neither determinately true, nor determinately false, nor determinately anything else—not even determinately vague if the case is sufficiently borderline.

It is also not quite correct to say that the coherence theorist is committed to rejecting bivalence. For one thing he might draw up his theory in such a way that the coherent system of beliefs was adequate to determine every statement as true or as false. He could do this either by making sure that the coherent system was large enough—equating it, for instance, with the system of *God's* beliefs—or alternatively by restricting what is to count as a statement so as to admit nothing that is not determined as true or as false by the coherent system. This latter would be the position of certain verificationists. What *is* true, though, is that unless he takes one of these steps he will not be entitled to assert that the principle of bivalence is true. Nothing within the coherent system will warrant the denial of the possibility that there are statements such that neither they nor their negations are determined as true by the system. Admittedly, as Blackburn points out, there might be some practical value in his deciding to treat the principle of bivalence *as though* it were true; deciding, that is, to treat every statement *as though* it must be either true or false, rather as a judge may decide that every putative contract must be treated as though the legal system determined it either as valid or as invalid. But this is not the same as holding the principle of bivalence to be true literally.<sup>11</sup>

If he is not entitled to assert that the principle of bivalence is true, it does not follow that he is entitled to conclude that it is false. This will only follow if it is assumed that the totality of truths is determinate, in the sense that for any statement it is a determinate matter whether it belongs to the totality or not. Given that assumption, if the system allows for

statements such that neither they nor their negations belong to it, these statements will be determinately neither true nor false, and consequently the principle of bivalence will itself be false. But it would be possible to deny the assumption, and some coherence theorists do: in their view the propositions that cohere with the system do *not* form a determinate totality. It would be natural to think that the totality of truths was indeterminate if one thought there was nothing more to truth than what we can recognize as true, and if one regarded it as somehow open-ended what we can recognize as true: as something that cannot be firmly delimited in advance of investigation. This is the line of thought taken by those who call themselves anti-realists, and as we shall see it was also taken by Kant.

Thus in fact there are three possibilities. (1) The coherent set may be such as to contain either  $p$  or not- $p$ , for every value of  $p$ ; the resultant system would be bivalent. (2) The coherent set may be such as to contain a specification, for each value of  $p$ , as to whether  $p$ , or not- $p$ , or neither  $p$  nor not- $p$ ; then the principle of bivalence would be false of this system. But (3) the coherent set may not form a determinate totality, so that for some values of  $p$  the status of  $p$  may be indeterminate. This possibility subdivides further. (3a) It may be that there are certain values of  $p$  to which the system determinately assigns the status 'neither true nor false'; in that case, again, the principle of bivalence would be false of the system. This, as I shall argue in the next chapter, is the position of Kant: although admittedly he does not use this terminology, in effect he regards the totality of truths about the world of appearances as being indeterminate, but also considers that so far as the world of appearances is concerned the principle of bivalence is determinately false. (3b) On the other hand it may be that there are *no* propositions to which the system determinately assigns the status 'neither true nor false'. In that case the principle of bivalence would not be determinately false of the system, but it would not be determinately true of it either. This is the position that Dummett and others call anti-realism. Anti-realists equate something's being determinately true, in virtue of the coherent system, with our being able to recognize its truth, and they equate something's being determinately false with our being able to recognize the truth of its negation; but they also do not think it possible to find any statement which is determinately neither true nor false. Dummett puts this by saying that the anti-

realist accepts the principle *tertium non datur*—the principle ‘that there can be no circumstances in which a statement can be recognised as being, irrevocably, neither true nor false’.<sup>12</sup>

It may seem less than obvious that anti-realism is a form of coherence theory of truth. Anti-realism is essentially a theory about meaning and understanding: it holds that to understand the meaning of a sentence is to understand the conditions under which it is warrantably assertible. These conditions may not be such as to ensure its truth, and indeed there are various types of statement—statements about the past, for example, or about other people’s mental states—for which conclusive verification is never available. This naturally gives rise to the thought that these statements must *have* truth-values independently of our verification, and that just because our verification can never be conclusive, there must be some truth of the matter lying for ever beyond our reach. But it is exactly this that the anti-realist most firmly denies. To say that the verification is never conclusive is to say that though our assertion of such a statement may be well warranted in the circumstances, our warrant for it is always defeasible; more evidence could always turn up which could render the assertion unjustifiable after all. There is no determinate totality of evidence which could settle the matter conclusively. The statement *has* no truth-conditions independent of our capacity to recognize it as true. To put it another way, he holds that there is nothing to its being true over and above its being recognizable as true by us.<sup>13</sup>

Anti-realists hold this because of their views about meaning. They consider that we can learn how to use words in assertoric sentences only by learning in what circumstances these sentences can justifiably be asserted, and it is only through the appropriate use of such sentences in such circumstances that our understanding of them can be exhibited. Since meanings can be learned and can be exhibited, to understand a sentence can be nothing more than to know the circumstances under which it can be justifiably asserted. Truth-conditions which we could never know about cannot enter into our understanding of a sentence, because there is no way in which we could either learn or manifest an understanding of them. Since the sentence is a sentence of our language it can therefore make no sense to suppose that it has truth-conditions of that kind at all, or that the statements we can use it to make possess a

truth-value independently of what we can discover. It is the realist view, to which anti-realism is opposed, that whenever a sentence is used to make a specific statement, that statement is true, or false, or enjoys whatever intermediate status there may be, regardless of whether we can know that. According to the anti-realist this idea of verification-transcendent truth is a myth. The idea of verification-transcendent truth is just the idea of truth that the correspondence theory makes use of. In rejecting it, the anti-realist adopts the alternative view, that truth is not independent of our capacity to find out about it, or in other words to have beliefs about it—beliefs that are warranted in their context.<sup>14</sup>

What does it mean to say that a belief is warranted in its context? The context may consist of other beliefs which support it, or of perceptual circumstances, or both. With many beliefs, like the belief that it is cold or that there is a table before me, it is natural to feel that it is the perceptual circumstances that warrant them. But for the anti-realist the fact that such-and-such perceptual circumstances obtain cannot itself be independent of our recognition of it, any more than any other fact can be independent of our recognition. Hence even where a belief is warranted by something perceptual, it is still in effect another belief that warrants it; and this means that we have on our hands a pure form of the coherence theory of truth. Beliefs must fit in appropriately with other beliefs which are themselves warranted in the same fashion, through coherence; and so far as it makes sense to talk of truth at all, truth is a matter of what we can in this fashion recognize as true.

Where anti-realists differ from many other coherence theorists is in not thinking of the set of truths as a determinate totality—as was observed above. ‘What we can recognize as true’ is indeterminate and open-ended. But their theory is none the less a coherence theory, in the sense explained. For what counts as true is determined by what we are able to discover. It may be objected that the anti-realist’s talk of discovery and recognition implies that there is something there to be discovered or recognized, independently of our discovering or recognizing it, and that would seem incompatible with the coherence theory. But care is required here. It is open to anti-realists to hold that there are plenty of truths which we have not yet recognized, provided they acknowledge that what makes them truths is that we are capable of recognizing them;<sup>15</sup> in this they are in the

same position as many other coherence theorists, who hold that in order to be true a statement does not have actually to be believed but only to fit with the system in the appropriate way. If, however, the suggestion is that there must be truths there to be discovered or recognized independently of our *capacity* to discover them or recognize them, then it is something anti-realists repudiate firmly, for as we have seen their account of meaning gives them no way to understand any such idea of truth.

Anti-realists may well say that they are not offering a theory of truth at all, but only a theory of warranted assertibility. If they do, however, we must ask what point they are seeking to make in saying it. If their point is that there is no determinate totality of truths, or that the principle of bivalence does not hold, then there is no need to disagree with them: if a 'theory of truth' is committed to bivalence and to there being a determinate totality of truths, then certainly their theory is not a theory of truth. But it remains the case that they are committing themselves to a 'theory of the nature of truth' in the sense in which I have been using these words, and to holding in particular that its nature consists in coherence. They might, perhaps, attempt to deny that any such claims about the nature of truth are warrantably assertible, on the grounds that they are too general and abstract. But if they consider themselves entitled to assert that no statement can have verification-transcendent truth-conditions, and if they also recognize that the aim of warranted assertion is to describe how things are, they should regard themselves as equally entitled to assert that how things are is not independent of how we can discover them to be.

In the last chapter I claimed that it was an attraction of certain forms of the coherence theory of truth that they could offer an answer to the extreme form of scepticism, by allowing us to rule out the possibility that our beliefs were radically false. In so far as the anti-realist regards our warranted assertions as being open-endedly defeasible, anti-realism cannot quite claim this advantage, because we can never be confident that circumstances will not arise which will defeat the ascriptions of truth-value that we have so far made. Typically, however, anti-realists do not regard *all* our assertions as being defeasible in this way; they hold that there are also types of statement that are susceptible of conclusive verification, and which we may therefore be able to know (indefeasibly)

to be true because we have verified them. They could therefore resist extreme scepticism over the truth of these statements: not even the *malin génie* could make them false, since it is in our recognition of it that their truth consists.<sup>16</sup>

Moreover, and perhaps more importantly, the anti-realist has a reply to the sceptic even where statements of the defeasible type are concerned. For although we have no assurance that our judgments in this area will not be defeated, we do have an assurance that they are justified, provided we can recognize that the circumstances obtain which warrant their assertion. It is part of what we learn when we learn the meanings of the words involved that such-and-such a statement is warrantably assertible—and therefore justified, although defeasibly—in such-and-such circumstances. And here again the *malin génie* could not be deceiving us; there is no possibility that this notion of justification is merely our notion, and matches nothing in reality. The reality of the matter is determined by how we think.

At any rate, it is clear that anti-realism is a version of the coherence theory of truth. I said I would conclude this chapter by looking at the relations between the coherence theory and idealism, so it is natural to start by asking whether anti-realism is a form of idealism. Here, of course, a lot depends on what idealism is. Like many ‘-ism’ words the term is variously used. In one sense idealism is the theory that material-object statements are reducible without loss of meaning to statements about mental states or Berkeleian ideas; clearly the anti-realist is not committed to idealism in that sense. In another sense idealism is sometimes said to be the theory that our minds create the world; but this graphic metaphor is more likely to be misleading than to be helpful. Dummett himself makes use of a different metaphor when he associates anti-realism with the picture ‘of objects springing into being in response to our probing’, and says: ‘We do not *make* the objects but must accept them as we find them ...; but they were not already there for our statements to be true or false of before we carried out the investigations which brought them into being.’<sup>17</sup> If that is not misleading it is only because it is obscure where it leads. To speak of us as creating the world suggests we have some choice in the matter, and that we make it in something like the same sense in which we make machines or write novels; these suggestions



are fairly obviously false, and neither anti-realism nor any other coherence theory is committed to them. Dummett's own metaphor suggests that before we investigate there is nothing there, hardly a welcome thesis and one that anti-realists are in any case not committed to, since like any sensible coherence theorists they regard a statement like 'There were dinosaurs before there were people' as making a well-warranted scientific claim. Dummett's metaphor also suggests that when we do investigate, the *nature* of what we find is independent of our cognitive capacities—a conclusion which as we have just seen the anti-realist is bound to repudiate.

In this area of philosophy (as in a number of others) metaphors are very tempting, but also dangerous; one should use them only when one is clear how they can be cashed out. If idealism is taken to be the thesis, not that minds or our minds create nature, but that how things are depends wholly upon some mind or minds, we have a more seriously discussable thesis; though it needs to be made clear that the dependence envisaged is logical and not causal, or the traditional account of the physical world as caused by an act of God's will would have to count as an idealist theory. If idealism is thought of in that way, its affinities with anti-realism, and with coherence theories of truth in general, become clear.

It seems clear, in fact, that not just anti-realists but any coherence theorists must be idealists in this sense. For they hold that truth consists in coherence with some system of beliefs; in the anti-realist version, that how things are is determined by our capacities to recognize them as such, i.e. to come to the belief that they are, in a context in which that belief is warranted (by its relation to our other warranted beliefs). In that case the character of reality is logically determined by certain beliefs. And it would not do, of course, to say that these beliefs might be the beliefs of no mind, for then they could not be distinguished from mind-independent facts about the world, logically independent of what is thought about them, and that is just what the coherence theorist rejects.

Must every idealist be a coherence theorist? One suggestion might be that the idealist could hold that truth consists in some relationship between beliefs, but claim that the relationship was something other than *coherence*. As I have said, though, the term 'coherence' is used in various different ways in different versions of the theory, and it might not seem unreasonable to count any truth-constituting relationship between beliefs

as a form of coherence. If this suggestion is ruled out it might appear that idealists do have to hold a coherence theory of truth; but in fact they do not, for there is another possibility. For one might hold that although the nature of the world entirely depends upon mind, it depends upon some other aspect of mind than the propositions it accepts. On this view what would determine the truth would not be beliefs, but mental states of some other kind or kinds. They might include, for example, perceptual presentations; Berkeley's account of the material world is idealistic in the sense we are concerned with, but his 'ideas' would seem to be perceptual presentations, or copies of them, rather than beliefs. F. H. Bradley, moreover, who is often described as a coherence theorist—and who often writes like one—is in the last resort not one, because what determines reality for him is ultimately not belief but feeling.<sup>18</sup>

Thus not every idealist need be a coherence theorist, but every coherence theorist must (if he is to be consistent) be an idealist. It may seem a surprising conclusion that every coherence theorist must be an idealist. The coherence theorist may be verificationistically minded, as indeed anti-realists are: and verificationistically minded people have often expressed doubts about the meaningfulness of metaphysical theses like idealism, on the grounds that they cannot be verified or falsified. However the conclusion does hold, though only because 'idealism' is being interpreted in a comparatively harmless way. It is in fact formulated in such a way as to be entailed by the coherence theory, given that beliefs depend on the mind or minds that have them. Hence anyone prepared to subscribe to the coherence theory must be prepared to subscribe to idealism as so formulated; the coherence theory could not be verifiable without idealism being verifiable also. It may be thought to be perverse to formulate idealism in this fashion. Certainly, as I have said, the term is often used in other ways. But it represents a radical enough thesis in this form, and as I hope will become clear in what follows it is in essentially this way that many of those who have called themselves idealists—such as Kant, Fichte, and Bradley—have understood their own positions.

### III

So the coherence theory of truth does not fall to the standard objections, and whether it is ultimately defensible is an important question. If it is

not, antirealism fails, and so do a collection of metaphysical theories that can be called idealist. To assess whether the coherence theory is defensible, we must bear in mind that the coherence theory is a theory of the nature of truth, and we must be clear about what that entails. Truth is a property of beliefs or propositions, and an account of the nature of truth tells us what that property consists in, just as an account of the nature of heat tells us what that property, heat, consists in. It does not just tell us that something is true *if and only if* certain circumstances obtain—if and only if it has some other concomitant property, for example. Nor does it just tell us that something is true *necessarily if and only if* certain circumstances obtain. It tells us what truth is. If there is a necessary equivalence between something's being hot and its having a certain mean molecular kinetic energy, that doesn't show that this is what heat consists in, for it may be necessarily the case that something has a size if and only if it has a shape, yet having a size doesn't consist in having a shape (or vice versa). Likewise, it may turn out to be necessarily the case that a certain type of mental event coincides with the occurrence of a certain type of physical event, but it remains a further question whether the mental state consists in the occurrence of the physical one.

We saw earlier that the traditional contrast between correspondence and coherence theories depends on this. The kind of correspondence theory that properly contrasts with the coherence theory is one that offers a rival account of the nature of truth. There is no problem for a coherence theorist in agreeing that a proposition is true if and only if it corresponds with the facts, and for the same reason there is no problem for a correspondence theorist in agreeing that a proposition is true if and only if it meets certain standards of coherence. Some people have gone further and held that reality must *necessarily* exhibit some high degree of coherence, perhaps the maximum degree, if that makes sense. That still does not make them adherents of a coherence theory of truth, unless their grounds for holding this are that truth itself consists in coherence. And the grounds have sometimes been quite different. For instance, people have thought for metaphysical or theological reasons that there is an independent reality that has to be intrinsically coherent, and that what makes a proposition true is its correspondence with that coherent reality. F. H. Bradley, for instance, often seems to hold this. To the extent that

he does, he subscribes not to a coherence theory of truth but to a kind of correspondence theory (Bradley, 1893, 135–161; Walker, 1998, 98–108). As the example of Bradley illustrates, though, one needs to be particularly careful here, because some writers commonly said to hold a coherence theory of truth have either not really done so at all or else been unclear in their own minds whether truth consists in coherence or rather in matching a reality that is necessarily coherent.

As theories of the nature of truth, the coherence theory holds that truth consists in coherence with some set of beliefs, while the correspondence theory holds it consists in matching a reality that is independent of anything that may be believed about it. Some versions of the correspondence theory make strong claims about the nature of the match, seeking to give informative accounts of the correspondence relation and of the ontology of facts. Austin's theory is of this kind, as is Wittgenstein's in the *Tractatus*. But one can be a correspondence theorist without making claims of that sort. Mackie's theory of "simple truth," recently defended by Alston, holds just that for a proposition to be true is for things to be as it says they are, and this is a correspondence theory provided it is understood that things are as they are independently of anything that may be believed about them. (Mackie and Alston both do understand it in this way: Mackie 1973, 17–63; Alston 1996, 5–64.) Just as it seems rather obvious that a true proposition (or belief or judgement) is distinct from whatever it is that makes it true, so it seems clear that there must be some relationship in virtue of which it is true. The correspondence theory says it is a relationship between the proposition and a reality independent of beliefs about it, while the point of the coherence theory is to hold that it is a relationship between the proposition and a reality determined by a coherent system of beliefs.

One could try avoiding either theory by saying that something is wrong with such a broad question about the nature of truth; all that one can properly do is to talk about the conditions for the truth, or the warranted assertibility, of particular propositions. Now it may be that the question is too broad, or too basic, to be susceptible of any very exciting answer, and that the best we can do is to say, "For a proposition to be true is for independent reality to be as it says," or something of the sort. That, though, is still an answer, even if not a very surprising one. Some

would say it is unintelligible, on the grounds that such matters transcend verification. But we have seen that to adopt this antirealist approach is to commit oneself to the coherence theory, and therefore to holding that the unexciting answer just offered is not unintelligible at all, but false. Others would concede that it is intelligible, and that its coherence rival is intelligible also, but they would say that the only way to cast light on the notion of truth is to provide a set of specific conditions for each proposition, or for each individual assertoric sentence of the language, along the lines of “‘Snow is white’ is true if and only if snow is white.” This line of thought leads to what are called semantic and deflationary theories of truth. That need not, however, concern us here. We have seen that coherence and correspondence theorists must both accept that there is a harmless sense in which true propositions are those that “correspond with the facts.” Equally, coherence and correspondence theorists can both debate the merits of the deflationary approach. This is because the deflationists and their opponents are trying to answer a different kind of question about truth from the question that is at issue between coherence and correspondence theorists. It is not a question about the nature of truth but a question about how to characterize the meaning of the predicate “is true” so as to exhibit the relation between “Snow is white” on the one hand and snow’s being white on the other.

In effect, then, if truth does not consist in coherence, it will have to consist in correspondence. However, we must also recall the distinction drawn earlier between pure and impure coherence theories. Both are theories about the nature of truth. But pure coherence theories are theories about the nature of all truths: they hold that truth consists in coherence with some set of beliefs. Impure coherence theories hold that there are some truths that consist in coherence, but also that there are others that do not: some things just are the case independently of the coherent system; some truths consist in correspondence.

Pure coherence theories are not ultimately defensible. They collapse into incoherence, as I shall show in a moment. But impure coherence theories can escape this difficulty quite easily. Two problems remain for impure coherence theories, but opinions will differ as to how serious these problems are. One has to do with the concept of truth itself. The concept appears to be univocal. Can it really be that there are two kinds

of truth, one consisting in coherence and applicable within a limited field, and the other not consisting in coherence? The other problem has to do with the reasons for adopting coherence theories in the first place. These are often reasons that would seem to apply globally if they apply at all. People are worried about what sort of match there could be between beliefs on the one hand and on the other a reality wholly independent of our beliefs and our ways of thinking about it. They are worried too about how we could ever *know* what such an independent reality is like or how we could ever succeed in making our words apply to it—a worry recently sharpened for us by Putnam (1983) but very clear in the work of idealists like Joachim (1906) as well as in the verificationists. If truth is cut loose from verification, scepticism seems to threaten, and perhaps meaninglessness. Such concerns do appear to apply quite generally, and not to one kind of truth rather than another.

### Why Pure Coherence Theories Fail

Coherence theories maintain not that truth consists in coherence with some set of propositions in the abstract but in coherence with some set of beliefs that are held, or perhaps some set of beliefs that would be held, in specifiable circumstances. It is this point that generates the difficulty for pure coherence theories. They are unavoidably committed to it. To suggest that truth consists in coherence with some set of propositions in the abstract would immediately open the theories to Russell's Bishop Stubbs objection. There are plenty of different sets of abstract propositions that can determine standards of coherence that a given proposition conforms to, and plenty of them no doubt include the proposition that Bishop Stubbs was hanged for murder. But by saying instead that truth consists in coherence with some set of beliefs that are or would be held, the pure form of coherence theory runs into a difficulty of its own. This is because it cannot accommodate the factuality of the claim that a particular belief is held, or would be held under the appropriate circumstances.

According to the pure coherence theory, the truth that belief *b* is actually held must itself consist in its own coherence within the system of beliefs. It cannot be a fact, independent of that system, that *b* is held. If it were, the truth that *b* is held would be a truth that did not consist in coherence, and we should no longer have a pure coherence theory. The

same applies if instead of the truth that *b* is actually held, we consider the truth that *b* would be held under the appropriate circumstances. From now on I shall just talk about the truth that “*b* is held,” but what I say can be applied equally to “*b* is actually held,” “*b* would be held under circumstances *c*,” or any other variant that the pure coherence theorist might adopt.

The truth that *b* is held, then, must consist in its coherence within the system, on any pure coherence theory. We now have two questions: what is the system, and what is it for *b* to cohere within that system? These questions are not distinct, for we saw that what determines the system is not a set of propositions in the abstract but a set of beliefs that are held, and this set must include beliefs about what constitutes coherence if it is to provide any standard of coherence. However the answers might be spelt out in detail, the truth that *b* coheres within the system clearly consists in the coherence with the system of “*b* coheres with the system.”

This promises a regress. But it is not there that the problem lies, for the regress is not fatal. It is just a version of the regress one will get on any theory of truth: if it is true that *p*, it is true that it is true that *p*, and it is true that it is true that it is true that *p*. The problem arises because the system itself is determined not by reference to a set of propositions in the abstract but by reference to beliefs that *are held*.

If we are not pure coherence theorists, there is no difficulty for us here. We can just recognize it as a fact that these beliefs are held—a fact that obtains in its own right. To the pure coherence theorist, this course is not open. It is a fact that *b* is held, no doubt, but what makes it a fact is the coherence of “*b* is held” with the system. That means, in effect, the coherence of “*b* is held” with the various beliefs that determine the system. We might call them *x*, *y*, *z*. These beliefs can determine the system only because they are themselves held. But what determines that they are held is just the coherence of “*x* is held,” “*y* is held,” “*z* is held” with the system itself.

This gives us a new and decisive objection, more sophisticated than the Bishop Stubbs objection, though in some ways like it. There may not be room for as many coherent systems as Russell envisaged, but there will be room for a great many. There will be room for a system that includes most of our usual standards of coherence but that also requires coher-

ence with the belief that Bishop Stubbs was hanged for murder. We can no longer repudiate this by saying that we have here only an arbitrary set of propositions, and not a belief that is held by the appropriate person or group. For “Bishop Stubbs was hanged for murder” will be an actual belief provided that “It is believed that Bishop Stubbs was hanged for murder” also belongs to the coherent set, since this is all that is required to make it a truth that the belief is held. The same point applies not just to beliefs about bishops but quite generally to whatever beliefs determine the standards of coherence and define the coherent system that supposedly constitutes the truth. This just means that there is nothing that determines truth. We sought to determine truth by reference to actual beliefs, beliefs that are held, but what beliefs are actual beliefs depends on whether “*b* is an actual belief” coheres with the system. So which beliefs are actual turns out to depend on the coherent system, which can itself be determined only by reference to some given set of actual beliefs.

Equally, of course, if someone puts forward a pure coherence theory according to which what matters is not just that these beliefs be held but that they be beliefs of type *t* and held by the members of group *g*, then the fact that these beliefs *are* of type *t* and held by the members of *g* will just consist in the *belief* that they are cohering with the other members of the set.

For this reason, no pure coherence theory is tenable. A tenable coherence theory will have to leave room for certain truths whose nature does not consist in coherence. These will *have* to include truths about the beliefs that define the system and determine coherence. Otherwise, the theory cannot get going.

### Correspondence Theories Survive Frege’s Objection

This objection to pure coherence theories may look similar to an objection Frege once made against any attempt to define truth, and which, if it worked, would equally be an objection against any attempt to say what truth consists in. A correspondence theory, for example, will say that truth consists in correspondence with facts. In that case, for *p* to be true is for *p* to correspond with the facts. But whether *p* corresponds with the facts or not is a question of whether it is *true* that *p* corresponds with the facts. So we attempted to give an account of what truth consists in only



to find ourselves making use of the concept of truth again, so that “what is defined must itself be presupposed” (Frege 1969, 139–140).

But Frege’s objection is not a good one. If truth consists in correspondence with the facts, then the truth of *p* will consist in its correspondence with the facts. If it does correspond with them, then evidently it will be true that it corresponds with them. But its correspondence with the facts does not consist in this. According to the correspondence theorist, its correspondence with the facts is just basic and does not consist in anything else. According to the correspondence theorist, the facts constitute an independent reality, and whether a proposition matches that reality is itself a factual matter. If *p* does match that reality, then equally “*p* corresponds with the facts” matches that reality, and thus so does “‘*p* corresponds with the facts’ corresponds with the facts.” But that is simply, once again, a regress that may be interesting but is hardly vicious.

The pure coherence theorist, on the other hand, is unable to make the analogous move, which would be to say that *p*’s coherence with the system is just basic and does not consist in anything else. For the pure coherence theorist is committed to a general account of reality in coherence terms. To treat it as basic that *p* coheres with the system would be to deny that the truth of “*p* coheres with the system” itself consists in coherence. The truth of “*p* coheres with the system” would consist in its simply being the case that *p* coheres with the system. It might be the case that “*p* coheres with the system” does also cohere with the system itself, but it would not be this that made it true.

The correspondence theorist, who treats it as basic that *p* corresponds with the facts, does not have to deny that the truth of “*p* corresponds with the facts” consists in its correspondence with the facts. There would be a problem with that only if some new and different set of facts had to be involved, and some new correspondence. But no new set of facts is involved. In the independently real world, *p* is the case—its being the case is a feature of that world. That by itself is enough to make the proposition “*p*” true, and it is what that proposition’s truth consists in. In making it true that *p*, it also makes it true that “*p*” corresponds with the facts, that “‘*p*’ corresponds with the facts” corresponds with the facts, and so on. (The thought that a new set of facts might be needed at each stage perhaps arises through confusing propositions with sentences.

The sentences “*p*” and “‘*p*’ corresponds with the facts” are certainly quite different, and one might think their relations to other things must also be different. But they express equivalent propositions.)

Correspondence theories may say less than people would like. If a correspondence theorist says just that for a proposition to be true is for things to be as it says they are in a reality that is independent of our beliefs about it, we may feel we are hardly being offered a *theory*. Yet we are at least being offered something substantive, for it constitutes a rejection of the coherence theory and involves a commitment to a reality that is independent of what we believe about it. We have already seen that it is right to reject pure coherence theories. What about impure coherence theories?

### Impure Coherence Theories

Coherence theorists have often been rather unclear about whether their coherence theories are intended to cover the nature of *all* truth or whether there may be some truths to which the coherence account does not apply. They are often quite ambivalent, for example, as to how experience is to be handled. Some, like Neurath (1931, 1932–1933), are clear that judgements reporting the immediate content of experience owe their truth to their coherence with the overall system of beliefs, and this is what any pure coherence theorist is committed to (as Neurath was well aware). Others, however, have been uneasy about going so far and have been inclined to give such judgements a special status. This special status is sometimes extended to all judgements reporting on one’s conscious mental content, which may be taken to include judgements about what beliefs one has. Someone who goes this far can certainly avoid the objection to pure coherence theories, for the resulting theory allows that there are truths about what we believe that do not consist in coherence.

That may well leave it unclear, however, that the theory still has the advantages that made the coherence theory seem desirable in the first place. If the point is to overcome the gap between our beliefs on the one hand and the world they are supposed to describe on the other, the theory can now help only if the judgements in this special class are *incorrigible*, in the sense that they cannot be falsely subscribed to. Many philosophers have indeed thought some or all such judgements to be

incorrigible, though it is far from clear that they were right to. If the point is to avoid a problem about knowledge—how we can know about an independent reality—it may still be plausible to say that this problem is removed by a coherence theory of this type, provided we think that what judgements of the special class report is epistemologically secure in a special way so that no such problem arises over them. Then we can say that truths about other matters are epistemologically reachable because they do not concern a reality independent of our beliefs but instead consist in coherence. Judgements of the special class could have that epistemological security through being incorrigible, but there might be other and more plausible ways, though again it is far from clear that they actually are as secure epistemologically as they have often been claimed to be. Again, if the motivation for a coherence theory is semantic, because meaning is taken to be tied to the method of verification, we get the same result: it may be felt that judgements in this special class are verifiable in a particularly immediate way, so that unlike other judgements, they do not have to be explicated in terms of coherence. But as before, this is certainly contestable.

Actually, coherence theorists have sometimes gone further than this without being very clear about it, for they have sometimes been prepared to give a special status to principles like the law of noncontradiction, and perhaps to basic principles of inference generally. They have taken these to be definitive of “coherence,” in that (for example) they have assumed it to be incoherent to hold both  $p$  and not  $p$ . In doing this, they have apparently taken such principles to be objectively right, so that their own truth seems to consist in matching some independent reality of a presumably rather Platonic kind. But nobody putting forward a global coherence theory of truth ought to be saying things of this kind, and those who have made these assumptions have not, I think, been conscious that this is what they were doing. There is a long and regrettable tradition in philosophy of taking such principles for granted without asking about their status and how they are known. What a pure coherence theorist ought to say, of course, is that it is the system itself that determines what counts as coherence. If it is a feature of the system that it requires coherence with a set of beliefs that include the law of noncontradiction and

our other basic principles of inference, then that system will rule out holding  $p$  and not  $p$  together, and that's all there is to it. The same thing can and should be said in the sort of coherence theory we have just been considering, which is not entirely pure but remains as inclusive as is possible compatibly with avoiding the objection to which all pure coherence theories fall.

The main problem with a theory that seeks to be as inclusive as that is the one already noticed: that its plausibility may be weakened to the extent that the judgements in the special class turn out to be just as problematic as any others, in whatever way it was that inclined us to sympathy with coherence theories in the first place. There is another problem too, though perhaps not a very serious one. An impure theory, even an only slightly impure theory like this, has to give two different accounts of what truth consists in: one that applies to judgements in the special class and one that applies to other judgements. For other judgements, truth consists in coherence; for those in the special class, it does not. It seems odd to be suggesting that truth bifurcates in this way. There would then be two different kinds of truth, and it would appear misleading to use the same word for both of them.

When I wrote the book from which the above extracts came, I thought this second problem was quite serious. However, if one had good reason to adopt the kind of impure coherence theory we have been talking about, by the same token one would have reason to dismiss our intuitions about the univocity of "truth" and "true." Moreover, if we consider impure coherence theories that are less global—impure coherence theories that offer an account of truth as coherence that is intended to apply only to some much more restricted range of judgements—it is far from clear that our intuitions do not actually support the conclusion that two different kinds of truth are involved. People sometimes put forward coherence theories of mathematical, moral, or modal truth, for example. It seems rather natural to think that there might be some difference in kind between mathematical truths and ordinary truths about the world. Many will think the same about moral truths and modal truths. No doubt different people will have different intuitions about matters like this, but that should only caution us against putting very much weight on intuitions (or on the ordinary usage that encapsulates and expresses them).

The real problem about coherence theories is whether there are good grounds for holding them. It often seems to people that there must be something incoherent about the idea of a coherence theory of truth, and we have seen that the idea of a *pure* coherence theory is indeed incoherent. There is, however, no reason to think there is anything incoherent about an impure coherence theory, and an impure coherence theory confined to a limited area may be quite plausible. What generally makes coherence theories plausible is a worry about how our judgements can correspond to a reality that is independent of us and of our ways of thinking. Someone who thinks that there is no particular difficulty about this in the case of ordinary matter-of-fact judgements may think that the difficulty is acute in mathematics, in morals, or in dealing with counterfactual conditionals. In fields such as these, then, impure coherence theories may be attractive. How attractive they are will largely depend on how good the arguments may be for thinking that there is more of a problem in these fields than elsewhere with construing truth as correspondence, but here we are on familiar ground. It is important, of course, to bear in mind that what is needed is to show that there is something *special* about these areas that warrants providing for them a special account of truth. Some of the arguments put forward here for regarding truth as coherence are simply variants of arguments for regarding truth as coherence in general. If it is argued, for example, that the notion of correspondence is problematic here, or that truth in these fields would transcend possible verification unless it consisted in coherence, we need to be particularly cautious, because, taken in their general form, such arguments would apply to truth quite generally and lead us to a pure coherence theory of truth. And that theory is incoherent. So it is important to be clear what it is that differentiates the field within which truth is said to be coherence.

Impure coherence theories may or may not prove helpful in dealing with certain problematic areas. But because they are impure, they require us to recognize that there are at least some propositions for which truth consists in correspondence. The advantage of a pure coherence theory was that it promised to remove the problems over correspondence. No pure coherence theory is tenable. To show that is not to make those problems disappear. We have to recognize that there really are difficulties

with the idea of correspondence with fact and over how we can succeed in knowing about a reality that is wholly independent of us and of our ways of thinking about it, and that there really are difficulties in seeing how we can refer to or meaningfully talk about such a reality. These difficulties cannot be conjured away. They constitute a serious philosophical problem that we cannot avoid. There is a good reason for being disappointed in a correspondence theory that says no more than that truth consists in correspondence with an independent reality, and it is that it leaves these problems untouched. It may serve to remind us, rightly, that correspondence theories do not need to saddle us with an ontology of facts or to provide us with a complete analysis of the correspondence relation. But how can we succeed in saying anything about a reality that is genuinely independent of us and of our ways of thinking about it? And how can we have beliefs about that independent reality that so regularly and reliably reflect how that reality is, since that reality is independent of us in this radical way? What can ensure that our beliefs match a genuine reality and not the illusions of Descartes's *malin génie*? Perhaps these classic questions are too large and too fundamental for most philosophers to expect them to be answered by a "theory of truth." They therefore turn aside to deal with problems that appear more tractable. But these are questions to which any serious thinking about truth inevitably leads us.<sup>19</sup>

## Notes

1. B. Russell, 'On the nature of truth', *Proceedings of the Aristotelian Society* VII (1906–7); mostly reprinted as 'The monistic theory of truth', in his *Philosophical Essays* (London: Longmans, Green, 1910); and M. Schlick, 'The foundation of knowledge', in A. J. Ayer (ed.), *Logical Positivism* (Glencoe: Free Press, 1959). Cf. also A. J. Ayer, 'Truth', in his *The Concept of a Person* (London: Macmillan, 1963).

2. N. Rescher, *The Coherence Theory of Truth* (Oxford University Press, 1973), chs. 1 and 2.

3. There has been much discussion, most of it rather confused, as to what the bearers of truth and falsehood are. In what follows I shall take it for granted that they are propositions or statements, and I shall use these two terms equivalently unless the context indicates otherwise. They form the objects of belief, and the content of an unambiguous assertoric utterance on a particular occasion of utter-

ance; two utterances express the same statement or proposition if they have the same sense and the same reference. A particular *sentence* may thus be used on different occasions to express different propositions, though a sentence as uttered assertorically by a given speaker at a given time will express one specific proposition unless it is in some way ambiguous. As I use the terms, therefore, the possibility of a proposition's being neither true nor false is not ruled out by stipulation.

4. Russell, 'On the nature of truth', in *Proc. Arist. Soc.* vii (1906–7), pp. 32 ff., = pp. 156 ff. of his *Philosophical Essays*. L. J. Cohen ('The coherence theory of truth', *Philosophical Studies* 34, 1978) has pointed out that this will not in fact be true if the requirement for coherence is made as strong as mutual entailment, and the system is also powerful enough to contain every proposition or its negation as a member of the coherent set; there can be at most one coherent system that meets those constraints. But the constraints are so strong that we must wonder whether any theory that demanded such stringent conditions for coherence could be of much interest. The systems of Bradley and Blanshard, which Cohen has in mind, are certainly said to be such that every true proposition entails every other, but it is not clear that they mean quite what we normally do by 'entails'; nor is it at all clear that they satisfy the other requirement, by containing either *p* or not-*p* within the coherent set, for all values of *p*. But see also Walker 1989 the end of Chapter v, on Hegel, p. 100.

5. Cf. S. W. Blackburn, *Spreading the Word* (Oxford University Press, 1984), chs. 6 and 7.

6. See chapter ix of Walker 1989.

7. H. Field, 'Realism and relativism', *Journal of Philosophy* LXXIX (1982), p. 556; cf. S. W. Blackburn, *Spreading the Word* (Oxford University Press, 1984), p. 246.

8. Blackburn, *op. cit.*, ch. 6, sect. 6.

9. Blackburn, 'Truth, realism, and the regulation of theory', *Midwest Studies in Philosophy* v (1980), p. 353.

10. M. A. E. Dummett, *Frege: Philosophy of Language* (London: Duckworth, 1973), p. 466.

11. Blackburn, 'Truth, realism, and the regulation of theory', sect. III; *Spreading the Word*, ch. 6, sect. 4. It might be suggested that in the cases Blackburn considers the principle of bivalence is 'warrantably assertible' on practical grounds, and that it would therefore have to be considered true by an anti-realist (see below); for anti-realists call true whatever is warrantably assertible (though they admit that such judgments may be defeasible). But this exploits an ambiguity in 'warrantably assertible'. The warrant for asserting the principle is a practical warrant, a warrant for treating it *as if* it were true, and it is not this kind of warranted assertibility that the anti-realist regards as truth.

12. Dummett, *Truth and Other Enigmas* (London: Duckworth, 1978), p. xxx.

13. Dummett, 'Truth', with Postscript, in *Truth and Other Enigmas*; see also the Preface to *Truth and Other Enigmas*, esp. pp. xxii f.

14. Ibid., also Dummett, *Elements of Intuitionism* (Oxford University Press, 1977), ch. 7.1; and his 'What is a theory of meaning? (II)' in G. Evans and J. McDowell (eds.), *Truth and Meaning* (Oxford University Press, 1976).

15. How they fill out 'capable of recognizing' may vary from theorist to theorist. The problems with it, of course, are similar to the problems of the verificationists in determining what is to be meant by 'verifiable in principle'.

16. Cf. Blackburn, *Spreading the Word*, pp. 249 f.

17. Dummett, 'Wittgenstein's philosophy of mathematics', in *Truth and Other Enigmas*, p. 185; cf. 'Truth', *ibid.*, p. 18. For a proper discussion of the inadequacy of the metaphor see C. Wright, *Wittgenstein on the Foundations of Mathematics* (London: Duckworth, 1980), ch. xi.

18. As I argue in later sections of Walker 1989 and in Walker 1998, what Bradley really thinks is that the world is a coherent whole, and he holds a correspondence theory of the nature of truth, which nonetheless (because of the coherence of the world) allows him to use coherence as a very effective test of correspondence. As well as R. Wollheim's book, *F. H. Bradley* (Harmondsworth: Penguin, 1959), I have found very useful the collection of articles by A. Manser and G. Stock (eds.), *The Philosophy of F. H. Bradley* (Oxford University Press, 1984), and in this context particularly those by D. Holdcroft and J. Bradley.

19. Sections I and II of this chapter are excerpted from Walker 1989.

## References

Alston, W. P. 1996. *A Realist Conception of Truth*. Ithaca: Cornell University Press.

Bradley, F. H. 1893. *Appearance and Reality*. London: Swan Sonnenschein.

Frege, G. 1969. *Nachgelassene Schriften*. Edited by H. Hermes, F. Kambartel, and F. Kaulbach. Hamburg: Felix Meiner. Translated by P. Long and R. White as *Frege: Posthumous Writings*. Oxford: Blackwell, 1979.

Joachim, H. H. 1906. *The Nature of Truth*. Oxford: Oxford University Press.

Mackie, J. L. 1973. *Truth, Probability, and Paradox*. Oxford: Oxford University Press.

Neurath, O. 1931. "Soziologie im Physikalismus." *Erkenntnis* 2: 393–431. Translated as "Sociology and Physicalism" in A. J. Ayer, ed., *Logical Positivism*, pp. 282–317. Glencoe, Ill.: Free Press, 1959. Also translated as "Sociology in the Framework of Physicalism" in Neurath's *Philosophical Papers 1913–46*, pp. 58–90. Dordrecht: Reidel, 1983.

Neurath, O. 1932–1933. "Protokolsätze." *Erkenntnis* 3: 204–214. Translated as "Protocol Sentences" in A. J. Ayer, ed., *Logical Positivism*, pp. 282–317. Glencoe, Ill.: Free Press, 1959. Also translated as "Protocol Statements" in Neurath's *Philosophical Papers 1913–46*, pp. 58–90. Dordrecht: Reidel, 1983.



Putnam, H. 1983. *Realism and Reason*. Cambridge: Cambridge University Press.

Walker, Ralph C. 1989. *The Coherence Theory of Truth*. London: Routledge Press.

Walker, Ralph C. 1998. "Bradley's Theory of Truth." In G. Stock, ed., *Appearance versus Reality*. Oxford: Oxford University Press.

---

## The Case for Coherence

Linda Martín Alcoff

“Perhaps nobody yet has been truthful enough about what ‘truthfulness’ is.”  
Nietzsche

Many philosophers today on *both* sides of the continental/analytic division advocate abandoning the question of the meaning of “truth.” Many analytic philosophers consider the meaning of “truth” to be a primitive or transparent concept and thus unsusceptible to any further elucidation, unlike other metaphysical concepts such as personal identity, God, and event. Some take it to be a question of language and argue that declarations of truth add no meaning beyond the substance of a claim. On the other side, many continental philosophers consider the making of truth claims to be the conversational equivalent of bomb-throwing, an attempt to shut down discussion by an absolute authorization of one’s own beliefs so as to make them immune to criticism. “If my claim is true, why do I even need to listen to the contrary claims of another?” these philosophers ask. Yet still other philosophers, on both sides, consider the problem of truth entirely specious or the product of a kind of category mistake such as when we conflate the language game of science with the language game of everyday objects.

Meanwhile, in the world outside of philosophy, the meaning of “truth” is experiencing a serious crisis. The effect of liberation movements both within and outside the United States, as well as the information glut produced by new technology, has been to magnify the quantity and quality of dissonance between claims and points of view to a level of complexity few can take in. An increasingly sophisticated citizenry is increasingly aware of the political filters through which “truth” is discovered, whether by

scientists dependent on grants from the private sector or politicians playing to the polls or pollsters, who themselves develop multiple-choice questions on morally complex problems. Scientific claims overtake each other so rapidly and social scientists disagree so much that the attribution of truth seems hopeless about all but the most mundane and unimportant matters. Contra Alan Sokal, it is not postmodern philosophy that has invited relativism but the postmodern conditions of life in which relativism seems sometimes to be the “truest” account of “truth.”

Against Wittgenstein, then, I do not think the problem of the meaning of “truth” is necessarily specious. The problem of whether to believe the physicist’s description of objects may well arise from a conflation of diverse language games, but the problem of whether to believe the latest scientific theories about what one should eat to be healthy, even well-confirmed theories, arises in practical life. Moreover, the problem of whether to trust Western-based scientific and general reasoning practices is a live issue for formerly colonized peoples the world over. As Malcolm X put it, “What is reason to the oppressor isn’t reason to the oppressed” (1969, 133). If anything is true, it is that every theory pronounced justified and true has not always been true, and that political investments have played a formative role too much of the time. Epistemic pretensions to be apolitical thus ring hollow today for many of us, and the meaning of “truth” is no longer, if it ever was, transparently clear. But to give up on the possibility of truth can only handicap political movements and obscure future solutions to present economic and environmental crises. Thus, I believe, we need an account of truth, but one that can make sense of *these* facts about truth, its history, and the means by which it is generated.

In this essay, I will argue that a coherence theory of truth has the potential to explain how realism can coexist with a political self-consciousness about human claims to know. On the analytic side of philosophy, the main problem for the coherence theory of truth is that it is interpreted as necessarily antirealist and, more recently, as tantamount to an epistemic concept of truth in which truth is cashed out solely in terms of epistemic concepts, as an extrapolation from our understanding of justification or warranted belief, and not in terms of the way the world is. If the coherence theory of truth is indeed an antirealist and epistemic

concept of truth, this would make it quite unattractive for anyone but committed idealists. On the continental side of philosophy, the main problems associated with coherence theories of truth are very different but no less serious. Coherence is taken to be impossible to achieve and even the wrong goal in and of itself insofar as it assumes that truth and understanding will always come from coherence rather than rupture, revolution, or a heteroglossia of conflicting and proliferating statements. Despite these serious and diverse problems, coherence theories of truth are also occasionally defended by philosophers from both traditions—notably, Putnam, Foucault, and Gadamer—but even in the writings of these philosophers the issues are not always laid out very clearly.

I will proceed by explaining the motivations behind the coherence theory of truth, and why this account is not necessarily antirealist, merely epistemic, or mistaken in its pursuit of coherence. My explanations here cannot be exhaustive because of space constraints; my hope is only to sow seeds of doubt about the implausibility of the coherence theory.

A typical formulation of coherentism about justification goes as follows: “A belief is justified to the extent to which the belief-set of which it is a member is coherent” (Dancy 1985, 116). What it means for a set of beliefs to be coherent is more variously defined. Some minimalist formulations of coherence require only simple consistency, while other, stronger versions require mutual entailment. The problem with the latter requirement is that it renders most actual belief sets incoherent and therefore unjustified, while the problem with the former is that it would force us into the position of accepting questionable or even fictional systems as justified beliefs if they only have internal consistency. A middle position that avoids these problems requires that the elements in a belief set be mutually explanatory. This involves symmetrical relations of support rather than the relations of logical dependence implied in the concept of mutual entailment. Explanatory support can be offered in a number of ways: by inference, correlation, analogy, or even similarity. It is widely acknowledged that the criterion of judgement most often used in deciding truth is precisely its coherence to existing knowledge, and this is so to such an extent that we often second-guess the perceptual evidence given by our own eyes when it radically contradicts what we believe we know to be true.

If a coherence theory of truth is adopted along with a coherence theory of justification, then it is held that truth itself represents a kind of coherence. This claim is best understood as motivated by the desire to transcend the usual binary division made between the world, on the one hand, and human knowledge, experience, interpretive schemas, and the like, on the other hand. Knowledge has most often been defined as a correspondence relation between two essentially dissimilar entities: a linguistic item and a bit of nature or a phenomenological experience, a mental entity and a corporeal one, a systematized set of propositions and a *Ding-an-sich*. Truth has been located at the intersection, as a bridge spanning the chasm between two “worlds” or as piercing an obstructive “veil.” But truth has been taken to mean a pure representation of the facts as they exist on the other side of the chasm, as they exist in a world that has been imaginatively drained of all human input. This kind of view was characterized by Dewey as the “spectator theory of knowledge”; Adorno called it, with more impatience, “peephole metaphysics.” When human knowledge and the world are imagined to be separated in this way, then any taint of human interpretation or concern must be removed from the world side if one wants to obtain truth. It is indeed as if we were reduced to looking at the world through a peephole and the role of epistemology were simply to make sure that the glass is spotlessly clean.

In the Kantian and Hegelian tradition, the first and most obvious problem with this binary picture is said to be that it is impossible for human beings to remove all traces of our engaged concern with the world and the substantive conceptual models and interpretive systems we use to make sense of the world. It is impossible for us mortals to attain a God’s eye view, in other words—a fact that (oddly) leads some to skepticism. Hegel rejected this conclusion because he believed that the binary picture implicated in the peephole view is just as wrong in its metaphysics as in its epistemology. We are not peering through a peephole at the world but are always already in its midst and concernfully engaged with it in multiple projects at particular sites, and knowledge is the outcome of sensuous human practices involving some aspect of this world within which we live. The coherence theory of truth begins from this metaphysical picture rather than the binary picture, and it thus offers what I think is best thought of as an immanent account of knowledge, against the transcen-

dental account of foundationalism. That is, for coherentism, knowledge is ultimately a product of phenomena that are immanent to human belief systems and practices, social organizations, and *lived* reality, whereas for foundationalism, if a belief is to count as knowledge, it must ultimately be able to establish some link to transcendent phenomena or something that is entirely extrinsic to human existence (i.e., the way the world would be if we had never existed). Where foundationalism ties justification to an external realm beyond beliefs and belief sets and understands truth as a relationship of a certain sort with this external realm, coherentism holds to an understanding of knowledge as emergent from immanent relationships in which there is never a pure or clear separation or noninvolvement between what we misleadingly distinguish as subject and object or human and world. As Foucault put it, truth is a thing of this world.

The coherence theory of truth understands truth as immanent to the domain of lived reality, rather than as completely transcendent of any human practice or context. It should be obvious, even though it often is not, that for this view what is true is not arbitrary or under the complete control of the knower, since coherentism takes us to be always already in the world, neither its masters nor blinded by our own concepts to such a degree that we can see nothing else. Truth is not a human construct, reflective only of facts about human beings. Rather, it is an event that occurs in the context of a mediated reality.

To obtain an initial intuitive grasp of this view, a view which is often counterintuitive to those who are philosophically trained especially in the Anglo-American traditions, consider the analogy Gadamer draws between the ontology of art and the ontology of truth. David Linge helpfully explains Gadamer's views on ontology as follows: "Like the game, the text or art work lives in its presentations. They are not alien or secondary to it but are its very being, as possibilities that flow from it. . . . The variety of performances or interpretations are not simply subjective variations of a meaning locked in subjectivity, but belong instead to the ontological possibility of the work. Thus there is no canonical interpretation of a text or art work; rather they stand open to ever new comprehensions" (Linge 1976, xxv–xxvi). Gadamer himself suggests that such a view can apply to the situation of human understanding in general principally because linguisticity is a universal and irreducible ontological

feature of the human relationship to the world (Gadamer 1991, 476). Inquiry is in fact both linguistic and historically embedded, and although this fact will have a different degree of impact on different projects of inquiry, depending on whether the objects of knowledge are natural phenomena or human activity, there remains a ubiquitous impact. Given these realities, truth should be understood as an emergent event of practical engagements rather than a set of intrinsic properties. The best conception of truth is not going to be a correspondence relation but the achievement of coherence among the multiple and diverse elements involved in the process or flow of knowing practices. When the event of knowing comes together, when a harmony is achieved, though always temporary and local, truth occurs. The phenomenological experience we have of the disclosure of truth, its revelation, occurs alongside an experience of the pieces falling into place, or of the pattern of the whole emerging against the backdrop of what had previously been meaningless atoms of disparate beliefs. Falsity, the opposite of truth, is always experienced as what is senseless, what gives the lie to all we know and experience, what is incapable of being sustained because of its very incoherence. We can sometimes understand how others could be led astray by certain falsehoods, or how we ourselves at an earlier date could have believed them, but now, given the knowledge we have at this point, we find it impossible to seriously consider them without risking our sense of ourselves as sane.<sup>1</sup>

Truth is best understood as indexed to a set of specifics, which include not only what we can see at a given time and place but also where our thinking is at any given moment, as well as the relevant features of reality. This makes truth both plural and changeable, since it is relative to a context richly conceived. But it does not make truth arbitrary or subjective: given sufficiently specifiable contextual ingredients, we can objectively determine truth, in some cases perhaps even using a deductive-nomological method. And the so-called subjective elements—the interpretive schema of knowers, their horizons of understanding, the historically specific episteme—are never sufficient to establish truth. Truth becomes apparent when beliefs and practices cohere within a lived reality.

Truth talk is not merely empty talk; it is a form of discursive practice with associated effects. It is embedded within a lived corporeal context,

and not merely an ethereal linguistic realm separated from bodies, practices, and material reality. Truth claims are about that whole lived reality: they refer to it, intervene in it, represent it. To eliminate any analysis or articulation of the ontological dimension of truth serves only to conceal from examination these relationships between truth claims and reality. Truth claims are claims about the nature of human life, about experience, and about our natural environment: the ontology of truth is the explication of the meaning, contours, and limits of that “about.” Some post-modernists deny the ontological dimension of truth because they believe truth claims are about a constructed reality. Hence they say that the point is to negotiate the features of this constructed reality, not simply to represent it. This mistakenly implies that we have the ability to negotiate the features of our lived experience, which is, of course, only partly true. We can affect the meanings of events and the intelligibility of experience, we can alter practices and even our physical surroundings, but we cannot “interpret away” death, human suffering, and the hole in the ozone layer, or render such things meaningless. They constrain the reach of our interpretive constructions if only by demanding to be interpreted, to be given a signification and a meaning in some form, to be included in our account of the real. Certainly there are phenomena and events we cannot explain and perhaps cannot adequately represent, but these descriptive and explanatory limitations are not existential boundaries and thus cannot dictate the scope of truth. Survivors of war may have no words to express their experience of war, no conceptual categories to make sense of it or to represent it, yet it permeates their lived reality nonetheless. There may be similar events that human beings cannot describe but can only witness. We cannot allow language to circumscribe ontology, nor can we replace ontology with language, without erasing significant parts of lived experience.

Every discourse makes truth claims, assumes (and uses) evaluative criteria for plausible and implausible claims, and seeks some reference to some elements (however variously conceived) outside of itself. These criteria, claims, and attempts at reference need theoretical analysis, clear articulation, and epistemological criticism. Powerful discourses are powerful because they resonate in us and connect with other discourses, practices, or experiences, because they help us to make sense of some-



thing we have already experienced, or because they are reinforced by other powerful ideas. In short, discourses derive their power from their coherence relations and their supportive connections with other discourses, experiences, or practices. I find an idea or explanation compelling because it makes sense of other things I already believe I know. Alternatively, I resist an idea or explanation when it is counterintuitive, that is, when it conflicts with too many other things I believe I know. Seen in this way, an acknowledgement of the epistemic importance of coherence relations does not strain our intuitions but makes it possible to incorporate our intuitions within an account of real knowing.

The coherence theory of truth is aligned with an immanent form of realism, a realism without the transcendental evocations of a realm beyond human cognition or interference. An immanent realism eschews the Cartesian bifurcations between “man” and world, culture and nature, mind and reality. Such an ontology of binarisms is a social construct that does not conform to the phenomenological experience of living or to the realities of scientific practice (Rouse 1987, chapter 2). The concept of a thing in itself is, after all, just a concept. The fact that we cannot apprehend a thing in itself does not automatically mean we lose out on a piece of reality; it means only that a particular concept has been found not to fit reality. What is much more real than a conceptual thing in itself is the lived world we share—a world of complexity, ambiguity, and richness that exceeds simple dualism.

It remains the case that dualism provided a powerful explanation for false belief. If reality is separated from the mind, the mind can represent that reality either accurately or inaccurately. Changing justified beliefs can be explained without involving ontological changes in reality. False beliefs and dreams need no counterpart or reference point in the world if the mind is essentially autonomous.

How to characterize changed beliefs is an issue that arises not only within the context of philosophical reflection based on modernist assumptions but also within the context of everyday knowing. After having learned new, disturbing facts about my ex-husband, do I reassess the “reality” of our marriage entirely? Was my previous happiness simply the product of a false belief? Whom was I married to: the man as I know him to be now or the man I thought him to be then? Or to use examples

that do not involve complicated issues of personal identity, I used to believe that margarine is better for your health than butter and that sugar intake is linked to hyperactivity. Recent studies have convinced me otherwise, though I wonder what the next studies will suggest. How would a coherentist account of truth characterize such changes? A non-bifurcated realism seems to require that changes in justified belief imply changes in reality itself. How would a coherence account of truth explain false belief?

The stock-in-trade examples regularly used in epistemology and philosophy of science, such as changed beliefs about the shape of the earth, phlogiston, ether, or the make up of oxygen, are relatively easy to explain. The claims of natural science involve complicated inferences and large-scale theories, far removed from immediate experiences including direct observation and sensation. When a claim is especially theory-laden, it is intuitively obvious that there is no simple fact of the matter. To insist that electrons must either exist or not exist is to transport beyond their realm of application the practical rules for everyday discourse about observable items (as when I say to my son, for example, "Either there is a girl in your room or there is not!"). Putnam's internal realism can account for scientific changes quite easily by indexing claims to conceptual paradigms or research programs, which have their own sets of categories and posited entities. Scientific ontologies are internal to models of reality.

But surely it would be unnecessarily purist to say that such claims within science therefore cannot claim to be true. This would be to lapse back to a dualist assumption that electrons must be entirely a human projection if we cannot verify their existence in any simple or direct manner. If we think of the ontology of truth in more complicated ways than simplistic one-on-one correspondences, it is possible to account for the actual sorts of changes that routinely occur in scientific explanation, which rarely take the form of "*p* and then not *p*" but more often seem to be something like "*p, q, r, s, t, u, v*" and then "*p, q, r, s, t, z.*"

Because of how variables hang together, rarely if ever capable of being pulled apart, changes in scientific belief are not well represented as simple negations. Thus correspondence theorists need not feel compelled to say that belief *a* was false and now belief *b* is true, though with some nervous insecurity about belief *b*'s likely longevity. Nor will it ever be necessary

to claim, as some thought coherence theorists might, that belief *a* was true at time  $t_1$  but belief *b* is now true at time  $t_2$ . We can account for changes in scientific beliefs more easily by offering partial, complex reports in which it may make the most sense to say that prior, discarded theories contained some truths. This is just to say, of course, that we can take advantage of the notion of scientific knowledge as a progressive accumulation to account for changes in a way that does not entail simple negations. Such theories need not involve the claim that science's accumulation of knowledge about reality represents an increase in the percentage of science that corresponds to a transcendental world. Rather, partial changes that involve accumulations can refer to improved practices, greater explanatory reach, and other advantages that refer to the goals set by the research program rather than to a transcendental concept of reality.

To return to the examples I raised earlier, which represent the more realistic problems one might encounter in the process of living, when one radically changes beliefs, either about other people, states of affairs, or even one's own history and character, how should this be characterized? Correspondence accounts would seem to have no trouble. They would simply say that my ex-husband was such as he was all along, that my mistaken belief that he was my true soul mate was based on lies and mistaken beliefs, and was not true. Either sugar and butter are harmful to physical well-being, or they are not.

Coherentism has a decided advantage here. If truth refers to a constellation of elements, then a change in belief occasioned by an increase or alteration in the relevant constellation is not a simple negation but an *altered truth*. What was the character of my ex-husband and the nature of our marriage? I was happy for several years, we developed a strong degree of emotional intimacy and mutual support, and my subsequent revelations about him can never completely change that history. But my assessment of him and of our relationship lacked some important elements, and the coherence of that assessment did not survive my enlarged and altered reconfiguration of knowledge. It is not simply that he changed or that I changed, though of course we both did. It is that the truth about my lived reality and even our shared lived reality changed.

Thus the belief I had at time  $t_1$  remains true despite the fact that now, at time  $t_2$ , a different belief referring to the same person is true.

But the two beliefs are not simple contradictories. They do not refer to or involve precisely the same constellations of elements. Nor are they equal, with no standard of assessment to distinguish between their relative validities. Ordinarily, unless there exist strong reasons to do otherwise, we privilege later beliefs based on fuller experience. However, we also ordinarily forgive earlier beliefs, recognizing their temporal situatedness. It is true that I might say, "How could I have been so blind?" Radical changes of belief generally prompt some self-examination. So I might ask what weakness in my cognitive capacities and character produced such apparent blindness, though instead of using those terms I would more likely ask, "Why was I such an idiot?" But unless self-examination leads me to decide that I willfully deceived myself, avoided clear indications, and created my own dream world of happiness, I am likely to accept the earlier belief about my ex-husband as true in part. I will include such beliefs as that my ex-husband had some positive qualities, that we had some genuine happiness together, but that there was more to the story than I knew at the time. Thus, because the beliefs are *about* different things (including myself), I can accept both earlier and later beliefs as true without creating metaphysical incoherence or accepting an outright contradiction or succumbing to a dysfunctional relativism. We do it all the time.

In regard to the delights of sugar and butter, part of what has changed over the years is the health science's theoretical orientations, for example, toward holism and away from the assumption that diseases have single causes, and toward a fuller recognition of the significance of individual physical differences and differences of lifestyle. Whether butter is dangerous to your health depends in part on the kind of life you lead and your genetic inheritance. It also depends on what ingredients are going into margarine this season. Of course, it is true that new studies sometimes contradict old studies, even when the presuppositions remain stable. In this case, from a new constellation of elements a new truth emerges, arguably better than the first because it is based on a more extensive constellation. We might even sometimes want to say that a prior claim was simply false, though every changed conclusion need not

elicit this explanation, and to make charges of false belief, we need not have recourse to a transcendental ontology. In fact, given the complexity of life, of science, and of human belief, the reference to a transcendental ontology might create more conceptual problems than it can solve. One would think that most philosophers of science would be prepared to acknowledge this after the exhaustive problems they have wrestled with throughout the twentieth century to maintain scientific realism in the face of scientific developments.

Putnam offers us another image that may help provide intuitive access to the coherence theory and that can also address the claim that coherentism is antirealist. The antifoundationalist philosopher of science Otto Neurath offered a famous image of science as forever having to reconstruct a boat while it is out to sea. This image has been widely interpreted as teaching that

in place of ultimate justification there can only be spot checks, and that these legitimating procedures themselves needed legitimation which in turn cannot be foundational either. Scientific knowledge is a communal project that has to hold itself in place. . . . Rather than chase the pipe-dream of apodictic grounds for human reason, Neurath urged us to explore the embedding of reason in what is not reason, to render intelligible the actual workings of reason, and, where possible, expose them to conscious intervention. (Cartwright et al. 1996, 93)

Neurath's metaphor reminds us that science is developed within a context of engaged interest and involvement rather than one of detached, affectless objectivity. It also nicely avoids the specter of a Berkeleyan idealism or subjectivism by casting us out to sea, where clearly we are bound to need reliable knowledge in order to stay afloat. Putnam borrows this image but helpfully revises it in the following way:

I would put ethics, philosophy, in fact the whole culture, in the boat, and not just "science," for I believe the parts of the culture are interdependent. And second, my image is not of a single boat but of a *fleet* of boats. . . . People are passing supplies and tools from one boat to another and shouting advice and encouragement (or discouragement) to each other. Finally, people sometimes decide they don't like the boat they're in and move to a different boat altogether. (And sometimes a boat sinks or is abandoned.) It's all a bit chaotic; but since it is a fleet, no one is ever totally out of signaling distance from all the other boats. (Putnam 1983, 204)

Putnam's revision points out the multiplicity of projects and paradigms used within human inquiry, and it is also helpful in explaining the rela-

tionship of the coherence theory of truth to relativism, an association many consider damning and proof of the failure of coherentism. Relativism first of all needs to be unpacked, and its various versions differentiated. Putnam rejects what he calls “unbridled relativism,” which would amount to the view that any substance whatsoever might be used to construct a boat and keep it afloat, an obvious implausibility. But a coherence theory of truth must commit one to some degree of relativism, since coherence, like the building of boats, can almost always be realized in more than one way, as Putnam acknowledges:

Why should there not sometimes be equally coherent but incompatible conceptual schemes which fit our experiential beliefs equally well? If truth is not (unique) correspondence then the possibility of a certain pluralism is opened up. (Putnam 1981, 73; see also Putnam 1983, 10)

Putnam himself espouses conceptual relativism and opposes ontological absolutism, but by conceptual relativism he does not mean that all theories are approximations or limited to a given subject’s particular standards. Nor does he mean that rationality is exhausted by current cultural norms or that every claim is potentially true relative to some model. Nor does Putnam understand conceptual relativism as entailing that we should (or can) relinquish the goal of decreasing conflict across diverse explanations or the goal of maximizing theoretical unity in the sciences, goals that have pragmatic motivations. Conceptual relativism amounts to the belief that, as Gadamer says of art, the variety of possible conceptual models and ontologies can be understood as expressions of the multiple possibilities inherent in the world itself. It is not that we fail to attain the truth in all cases of multiple interpretation, but that the truth of the world itself is closer to a kaleidoscope without borders than a multiple choice test where each question has only one right answer.

Why is there debate over whether such a view can be accommodated within the rather broad understanding philosophers now have of realism? Realism has hardly enjoyed a uniform or continuous meaning in the history of philosophy, but today it is generally taken as the view that reality would exist even if human beings did not *and* that we can know reality. Neither of these claims entail ontological absolutism. Moreover, embracing ontological pluralism does not commit one to the idealist notion that reality is a reflection or construction of the human mind, and it is in

fact compatible with more than one metaphysical system. For instance, ontological pluralism could be explained by the fact that reality is rich enough to admit of more than one conceptual or ontological expression. But even more important, in the tradition of philosophical hermeneutics developed by Heidegger or Gadamer and in some versions of neo-Hegelianism and phenomenology, the world we know admits of more than one ontology not because we can constitute it in various ways (which gives the human subject too much control and centrality), but because we are engaged with and related to the world in a variety of specific locations and with a variety of specific projects. To avoid the metaphysical picture conjured up by the phrase “constituted by mind,” the phenomenological tradition suggests that a more accurate characterization is to say that “the world appears to me.” This deflates the epistemic agency of human knowers in a way that is phenomenologically more apt to lived experience: in this appearing to we can be quite passive. We are not always or even often the masters of our epistemic domain, and in this formulation, it is the world that acts, but it acts toward me or in my direction. Thus, the “appearing to” language connotes a relational experience; knowledge occurs in an interaction, which is not well represented either as a constituting act nor as a spectatorial exercise. Such a realism seems more realistic.

Let me return to the problem of relativism, which remains for even a realist account if it holds to ontological pluralism. Paradoxically, Putnam, borrowing a familiar argument from Nelson Goodman, argues that we can avoid an “unbridled relativism” when we come to understand that there is no neutral position from which to judge competing claims. This may seem paradoxical if one believes that only such a neutral position could limit relativism. But the sort of unbridled relativism Putnam rejects is a relativism that would relativize *all* things, including even potential truth claims, and this requires a neutral position from which to make the general claim that all possible truth claims are relative. From the perspective of any other, nonneutral position, i.e., any particular conception of the world and of our place within it, all competing claims will never appear equally true but will be evaluated in light of the particular epistemic location one occupies. Putnam distinguishes himself from Goodman’s view in that “in the final instance” he strongly affirms, not

the claim that all views are relative, but that our versions are superior to others “*by our lights*, not by some inconceivable neutral standard” (Putnam 1983, 168). Whereas Goodman highlights the fact that, strictly speaking, relativism follows from the lack of a neutral standard, Putnam highlights the fact that precisely because no one is neutral, *we* (for *any* given “we”) cannot embrace relativism except of a rather empty metaphysical kind.

This is not to say that the problem of relativism arises only among metaphilosophers. Aside from the ubiquitous undergraduate examples, relativism is often invoked to address cross-cultural systems of belief. Theorists like C. G. Jung, Claude Levi-Strauss, and Joseph Campbell advanced universal projects that attempted to sift through cultural differences for the sedimented basic unities underneath, but these have become less plausible and less influential as Anglo-Europeans grow more knowledgeable about other cultures. The waning influence of these universalist systems of interpretation has in turn given rise to a relativist response, in which cognitive differences are said to be explained by culture or discourse, beyond which there is no appeal. Though this might seem to increase the likelihood of violent intercultural conflict, since relativism gives up the hope of resolving differences through rational discussion based on shared beliefs and values, it is often believed to lead actually to more harmony, because the adoption of cultural relativism would undermine the West’s superiority complex by rendering any overall comparison impossible. Relativism counsels tolerance, it is believed, whereas nonrelativism engenders accusations of irrationality or willful malice and a dogmatic attachment to one’s own cultural prejudices. Coherentism, because it requires the possibility of comparative judgement, is then sometimes argued to be an impossible pursuit as well as politically retrograde.

Although I have many political disagreements with Putnam, in this area our views converge. The problem with the relativist path just outlined is that, in the guise of promoting mutual respect, it renders real respect impossible. Tolerance for another person’s or culture’s incomprehensible positions yields not respect but grudging noninterference. If I cannot truly understand another group’s view or why they support it, I cannot truly take it seriously; I am left to view the other group as a mere



“noise-maker,” as Putnam (via Wittgenstein) maintains, or as a curious species interesting primarily for its entertainment value. Anglo-Europeans have too often taken this *National Geographic* attitude toward other cultures, the view of them as offering interesting customs and attractive costumes for Western enjoyment but unable to engage in serious dialogue or to share decision-making power over any important issues such as nuclear weapons. Thus beneath the tolerance and curious interest always lies a profound sense of superiority. If I cannot grasp the plausibility of another group’s views, I cannot truly accord them the same status as my own. Such accordance as I might perfunctorily bestow if I considered it required by “logic” would be merely metaphilosophical, having no association with my cognitive or affective state and therefore unlikely to affect my practice in any way. My tolerance would be based on an abstract acceptance of cultural relativism, on my Kantian duty, and not on a genuine understanding of why the other group’s views deserve my respect. This argument entails not that I must be able *fully* to understand a different set of views before I can truly respect them but only that there must at least be some real doubt about the adequacy of my own different views, some specific reason to think the other group’s views might just have something to teach me, before my respect can be genuine. In the face of conflicting beliefs, coherentism counsels not simply that we ignore the conflict but that (a) we acknowledge that in some cases truths can be plural but also that (b) we seek ways to maximize coherence by striving for an understanding of the alternative view to see if harmony can be achieved.

Moreover, the actual situation we increasingly find ourselves in rarely allows for a noncommittal tolerance, and certainly not when the differences concern anything of importance. Societies are so interwoven today that few if any exist separate from others; we are increasingly interdependent and mutually influencing. What usually happens when we encounter difference is that we strive to make sense of it somehow, to understand it within its own terms or by making an analogy with something more familiar. And this is a better response than the relativist ideal of disinterested tolerance. It is mistakenly believed by some critics of the Enlightenment that striving to understand difference necessarily entails its assimilation to our own norms of rationality and morality, and thus

necessarily reproduces chauvinism. This effect may be common, but it is not inevitable. Absolute relativism produces a tolerance for other cultures, but it also removes the need to revise one's own beliefs and practices in light of their difference from others'. If one strives truly to be "open to the Other," as Gadamer put it, and if one is willing to revise one's own norms rather than holding them back as nonnegotiable, an understanding can be achieved that does not silence or erase the other's "otherness." This process requires no pretense of being completely open or of operating from a *tabula rasa* with no interpreting systems or prejudices. Gadamer does not counsel us simply to ape the other, accepting all we come across without reflection or critique. His coherentist hermeneutics brings all the elements of contrasting meanings and forejudgments into play, without privileging either our own or the other's. And again, this is entailed by the very respect we wish to show the other. The liberal attitude of treating others (especially others whom one feels superior toward) with such deference and delicacy that one never voices one's own views or criticisms is patronizing, condescending, and revealing of a lack of respect.

Note that the previous discussion employs both political and epistemic considerations in reaching its conclusions, thus conforming to the kind of epistemology that Foucault called for in his claim that knowledge and power are inextricable. My argument against total relativism, like Putnam's, holds that it is both implausible and undesirable. Coherentist epistemologies can accommodate this multilayered evaluative approach because there is no linear chain of inference, as in foundationalism, but a conglomeration of considerations brought to bear on new items of potential belief.

The reduction of truth to a wholly epistemic meaning encounters problems similar to those of "unbridled relativism." The concept of truth that would allow for a total relativism and the perfunctory respect of everyone's beliefs is one devoid of metaphysical content. It is one that sees truth as exactly like the ball game that Wittgenstein describes where no rules are necessary: there is nothing at stake, and only agreement is required, but even this can be partial. An epistemic concept of truth would conform to such a picture, since it defines truth entirely in terms of epistemic practice rather than the qualitative relation of the result to

the real. It says that truth is a concept based on our understanding of justification, with no addendum beyond this. If this were the case, why claim a concept of truth at all? Claims to know the truth are claims *about something*, and how this aboutness is understood, as well as how the something gets conceptualized, necessarily brings us beyond epistemology proper and into metaphysics. A coherence theory of truth makes a metaphysical claim about the something and the aboutness: it claims that truth emerges when there is a coherent constellation of elements in an immanent domain of lived reality, and that this constellation includes more than purely human elements, even though it always includes those as well. It claims that coherence is the best way to conceptualize truth, given this metaphysical account of the knowing process, since correspondence presumes the viability of separation and would render unintelligible a plurality of ontologies with equal epistemic merit and metaphysical status.<sup>2</sup> Some versions of a coherence theory of truth might amount to a merely epistemic account, but my argument is that (a) there is no necessity to this fact and (b) in the most persuasive versions of the theory, truth is said to be coherence more for metaphysical than for epistemic reasons. The only viable metaphysical account of the knowing process is one in which knowledge is the product of situated human engagements with the world, and this best conforms to a coherence account of truth.

A remaining concern is the question of why we should hang our hats on coherence as the primary criterion of truth. Aren't some truths incoherent with past beliefs? Don't our eyes sometimes present us with the truth even though they contradict many strongly held beliefs? Moreover, aren't ruptures and conflicts invigorating for the progress of inquiry, as Bachelard and Kuhn argue?

Although it is certainly the case that the drive to maximize comprehensive coherence is a drive toward the resolution of conflict, this does not require the elimination of difference. One needs to distinguish between differences and conflicts. Some alternatives are not necessarily in conflict or competition with one another: physicalistic and mentalistic explanations of the same phenomena can peacefully coexist, as Davidson has shown, without calling each other into question. These contrasting explanations offer contrary concepts and categories and different causal narratives to explain a given phenomena or event. But they need not

conflict, because each is doing something different, engaging in a different project.

Of course, it is certainly true that all of the major conflicts in our broad array of human knowledge claims are not capable of such peaceful coexistence. Marxist and neoclassical economic explanations, for example, are contestatory frames for analyzing market behavior; Mumia Abu-Jamal is either guilty or innocent of the murder charges against him according to current legal definitions. But what the above example concerning physicalistic and mentalistic explanations suggests is that the mandate toward resolution is not an *a priori* dictate flowing from coherence in the abstract. Rather, the need for resolutions is more realistically understood as contextual, arising from specific problems in specific contexts. The scientific community is certainly far from a consensus on the need to achieve a grand unified theory that would somehow incorporate all existing scientific knowledge in a way that resolves any theoretical contradictions. From within specific research programs, a totalizing synthesis is rarely necessary to achieve their goals or even to justify their conclusions. If we take as our epistemological starting point the situation of “real knowing,” the mandate to achieve a total synthesis does not follow even from methodologies that base their procedures of theory choice on coherence. Michael Williams (1991, 295) argues that any coherentist epistemology would by definition seek the resolution of conflicting knowledge into a progressively coherent system. It is true that most coherence theories require comprehensiveness, so that coherence cannot be achieved by some *ad hoc* elimination of troubling, anomalous elements. But this requirement does not imply the need for a grand unified theory of all theories: that is far more than what one would need to avoid *ad hoc* solutions to anomalies, even for broad paradigms or research programs. Williams’s approach overall is too abstractly idealizing: it does not match up very well with actual processes of justification. Few believe that we know nothing if we cannot justify our global knowledge, and no one, of course, has a total coherent set of knowledge (except maybe cult followers, but even here I suspect some necessary fragmentation). The coherence theory itself does not mandate a global perspective on the whole of knowledge that would allow us to characterize its epistemic contents in a unified way. Williams’s more specific

target is the assumption that such globalism is possible, but coherence itself is not bound to this assumption.

A coherence theory of truth, then, can be contextually applied. Contextual boundaries can be determined on the basis of what is most relevant for understanding a specific claim or theory in a particular project of inquiry. One has to make a case for how one wants to draw the map of contextual relevance, a case that can have explanatory value. It does make sense to take industrial development and the consequent pollution as relevant to the prevalence of a local cancer; it does not make sense to take into account the Milky Way.

At the same time, I want to concede that in “real” knowing, we do not generally rest content with conflicts that might cast doubt on significant beliefs we hold, or that produce a felt dissonance between one set of beliefs or practices we hold and another set. Intellectual movements of the last two centuries have been profoundly affected by the need felt by many people the world over to resolve the conflict perceived to exist between religious cosmologies and the impressive successes of a science, which apparently calls these into question. Because the drive toward coherence is part of our actual knowing practices, there is a tendency toward conserving prior beliefs. New paradigms are not easily adopted, even when the evidence seems clearly in their favor. These facts should not be reduced to psychological weaknesses or lack of epistemological justification; the drive to make new beliefs and old ones cohere is obviously an important part of every successful program of inquiry.

It is also clear, however, that maintaining coherence is not all that’s important or epistemically valuable. Epistemic virtue is generally not accorded to those persons or theories that achieve coherence “too easily,” by dismissing contradictory claims without sufficient argument or by simply ignoring contrary evidence. One way to characterize facile achievements of coherence is within the dictates of coherentism itself. In other words, such facile achievements are arguably superficial and inauthentic, and thus not real achievements.

The drive toward coherence is obviously not all that is important in the development of knowledge, and at least some versions of coherentism can accommodate this fact. Contradiction, negation, opposition, and

tension play a positive role in the development of knowledge. Constant unpredictability and a certain amount of chaos provide a structure or framework for coherent theories in the same way that our mortality serves both to structure and to confer value on life. Meaningfulness is all the sweeter when one realizes how fleeting and contingent it always is.

Achieving coherence is never fixed or stable, and its very instability works to guard against facile or premature declarations of its achievement. The inherent exclusions of presentistic theories, which prompt some postmodernists to declare themselves “antitheory,” need not lead us toward valorizing only or primarily negativity. Two lessons we can learn from Hegel are relevant here. First, it is not negation per se that moves our understanding forward, but what he calls *determinate negation*. Abstract or generalized negation, exemplified by global skepticism, leads nowhere, points toward nothing, and therefore cannot guide the movement of understanding anywhere. Thus, contra Barry Stroud, to entertain global skeptical doubts cannot improve our knowledge, because such doubts can by definition offer no specific alternative procedures or concepts.

On the other hand, determinate negation, which is specific, does provide a specific direction. What this shows is that the value of determinate negation depends on its relationship to affirmation. The particular direction that a determinate negation suggests is conditioned by the specific affirmation it has denied. Thus we have the dialectic, in which negation and affirmation have an interdependence that is substantive for particular cases but also general and abstract. This is what indeterminate negation lacks, as antiskeptical philosophers have long noted. If global skeptical doubts are generated without any specific reason, they cannot be addressed, and the result is that conversation must come to a complete standstill. For this reason Peirce argued that skepticism is not even a genuine doubt, since real doubts are felt, arise within experience, and emerge within the context of specific goals.

Thus I fully admit that negation is a critical ingredient in the development of knowledge, but only if it exists in dialectical interdependence with affirmation. This points to the second and corollary lesson from Hegel: that the value of coherence lies in its dialectical relationship with

opposition. There is no question of simply affirming one or the other. However, clearly the drive toward coherence is primary in the *active* process of knowing. Negation, doubts, and conflicting evidence inevitably arise in the course of experience. One does not need to mandate the inclusion of conflict; one needs mandate only that we not try to ignore or downplay it as it arises. Knowing does not seek negation, nor does it need to do so. The movement of understanding is a movement toward coherence, which will inevitably meet negation all along the way and achieve coherence only intermittently and temporarily. The value of coherence itself, our motivation to pursue it and our felt desire for it, emerges only because it is so elusive and unstable, because we live amidst dissonance and contradiction. Thus the argument for coherence is not an argument that coherence is all one needs but an argument that the need for coherence arises within a context of incoherence. Positing the choice as one between coherence and negation is therefore misleading.

Epistemologically, there is no way to get around the fact that claims of truth involve exclusion, control, and the repudiation of opponents. The sphere of truth cannot be made politically correct according to the latest formulations. But after all, even freedom requires a context, just as postmodernists argue that resistance requires (and is the natural by-product of) power. The process of achieving truth must involve open dialogic exchange without *arbitrary* exclusions, such as those based on sexism or racism, but it must also involve exclusions based on epistemic status. Hence truth is not necessarily on the side of oppression in its self-privileging or assumption of authority.

The drive for coherence provides structure and a criterion of adjudication. The attempt to harmonize difference will necessitate some exclusion and control, but let us remember that coherence is a limit concept, or a heuristic; there is no possibility that we will ever achieve a totalizing coherence that will remain stable and thus prove stultifying. Hence the danger that coherence imposes is perhaps not so dangerous. As Foucault well knew, it is uselessly utopian to pursue the complete elimination of all structures of power/knowledge. For both epistemic and political reasons, what is needed is a decentered form of structuring, which can destabilize authorities so that they cannot become authoritarian. Coherence epis-

temologies posit no self-justifying states, no indubitable bedrock, that would create a foundation beyond challenge. Where the ultimate criterion is coherence within a large constellation of elements in a temporally and spatially specific context, rather than a foundation purportedly linked to a truth outside of history, truth will always be temporary and unstable.

## Notes

Portions of this essay are excerpted and revised from *Real Knowing: New Versions of the Coherence Theory* (Ithaca: Cornell University Press, 1996).

1. We can consider claims that might be true or false, of course, and even attempt to believe them, to try them out, as it were. And we can imagine what it might be like to believe a falsehood, but we cannot cross over to a *sincere* acceptance of beliefs we hold to be clearly false without losing our confidence in all knowledge we have of the world.

2. In its deflationary formulation, correspondence is compatible with ontological pluralism, but this is such an empty account of truth as to be of no competition to the coherentist view. Moreover, the history of the development of the correspondence theory of truth shows it to have been motivated by and linked to foundationalism.

## References

- Cartwright, Nancy, Jordi Cat, Lola Fleck, and Thomas E. Uebel. 1996. *Otto Neurath: Philosophy between Science and Politics*. Cambridge: Cambridge University Press.
- Dancy, Jonathan. 1985. *Introduction to Contemporary Epistemology*. Oxford: Basil Blackwell.
- Gadamer, Hans-Georg. 1991. *Truth and Method*. Second revised edition. Translation revised by Joel Weinsheimer and Donald G. Marshall. New York: Crossroad Press.
- Linge, David. 1976. Introduction to Gadamer's *Philosophical Hermeneutics*. Berkeley: University of California Press.
- Lyotard, Jean-François. 1984. *The Postmodern Condition*. Translated by Geoff Bennington and Brian Massumi. Minneapolis: University of Minnesota Press.
- Putnam, Hilary. 1981. *Reason, Truth, and History*. New York: Cambridge University Press.
- Putnam, Hilary. 1983. *Realism and Reason*. Volume 3 of *Philosophical Papers*. New York: Cambridge University Press.



Rouse, Joseph. 1987. *Knowledge and Power: Toward a Political Philosophy of Science*. Ithaca: Cornell University Press.

Williams, Michael. 1991. *Unnatural Doubts: Epistemological Realism and the Basis of Scepticism*. Cambridge, Mass.: Basil Blackwell.

X, Malcolm. 1969. *The Speeches of Malcolm X at Harvard*. Edited by Archie Epps. New York: William Morrow and Company.

# III

---

## Pragmatism and Verificationism



---

# Pragmatism and Verificationism: Introduction

## Classical Pragmatism

Pragmatism is the most influential and important American contribution to philosophy. Through its most famous advocates—Charles Peirce, William James, and John Dewey—pragmatist ideas have affected every aspect of American life, from psychology to education to the arts. And central to pragmatism is a particular conception of truth, one that sees a close link between truth and human experience.

In the classic essay included here, the founder of the pragmatist movement, Charles Peirce, introduces what William James would later identify as the key pragmatic maxim. In Peirce's view, if we wish to know the meaning of our ideas, we should "consider what effects, that might conceivably have practical bearings, which we conceive the object of our conception to have. Then, our conception of these effects is the whole of our conception of the object." By "practical bearings," Peirce means those effects of the idea that have a bearing on our practice, our action. The meaning of an idea, in short, consists in its practical effects on human experience.

Peirce's principle has a number of dramatic effects itself. For example, if two ideas have the same practical consequences, they have the same meaning, while ideas without practical consequences are meaningless. Most important, it suggests to Peirce a particular definition of truth, namely "the opinion which is fated to be ultimately agreed to by all who investigate." This definition is epistemic; rather than saying that truth is a matter of agreement with reality, Peirce suggests that truth is determined by agreement among ourselves. But not merely any consensus will do;

it must be consensus at the end of exhaustive empirical investigation—scientific inquiry when *complete*. In short, Peirce's view seems to be that a judgment is true if and only if it is justified at the end of scientific inquiry.

The great psychologist William James often credited Peirce for his own ideas about truth. But there are significant differences between their positions (indeed, Peirce ultimately disavowed James's version of pragmatism). While James explicitly accepts Peirce's view of meaning, and therefore accepts that truth has much to do with practical experience, James stresses that what is of practical relevance is in part a matter of our interests and attitudes. For James more than Peirce, the "trail of the human serpent is over all." Yet James also insists that the pragmatist can concur with the correspondence intuition that true ideas agree with reality. The pragmatist, according to James, is not denying this platitude; rather, he is trying to explain the nature of this "agreement."

James seems to give several different explanations of "agreement" and truth. In a Peircean vein, he claims that an empirical judgment is true just when it is verifiable. He also says that a true idea leads or guides us in our dealings with reality, and that a true judgment is what is expedient to believe. If there is an overriding idea here, it seems to be that a judgment is true just when it continues to be useful over time in our interactions with reality. The truth of judgment therefore consists in its continuous practical use in our lives, but what this practical use consists in, James seems to suggest, depends in part on the judgment in question. In this sense, James's conception of truth seems pluralist in design.

James's and Peirce's epistemic theories of truth have traditionally encountered stiff opposition. A typical charge against James is that he is committed to a form of relativism. For if the truth of a judgment consists in its practical use over the long run, then we might wonder whether there are any objective facts that determine when a judgment has such a use. If there are, then it seems that truth is determined by something other than practical use. If there are not, then whether a judgment is true seems to be relative to differing understandings of what is practically important, useful, etc. With his more straightforward identification of "practical bearings" with what is empirically verifiable, Peirce may seem to have less of a problem with the question of relativism. But Peirce's theory faces its own difficulties. One is that his theory simply assumes

that there is such a thing as a complete investigation, or that such an idea even makes sense. Another problem is that it seems intuitive to think that there could be judgments that are true (or false) but that are never discovered to be so by any scientific investigation whatsoever. Indeed, as we'll see, this intuition proves troubling for any pragmatist or verificationist theory of truth.

### Dummett and Verificationist Antirealism

It is difficult to overestimate the impact of Michael Dummett's essay "Truth" on subsequent work on the topic. The early part of the century saw an explosion of interest in theories, like those of the pragmatists in America or the positivists of the Vienna Circle in Europe, which linked truth to verification. But by the middle of the century, interest in these theories had flagged, stemming in part from the increasing popularity of the deflationary view. Originally written in 1958–1959, Dummett's essay served not only to revive the idea that truth may consist in verification; it introduced his influential idea that the debate over realism is at heart a debate over the proper way to analyze the concepts of truth and meaning. Indeed, Hilary Putnam would later credit Dummett for waking him from the dogmatic slumber of his metaphysical realism.

Dummett compares the concept of truth to the concept of winning a game. Just as it is part of the concept of winning a game that one plays a game to win, so it is part of our concept of truth that we aim to make true statements. Dummett parlays this insight into an objection against deflationary views of truth. A deflationary view, Dummett suggests, tells us when our statements are true by telling us that a statement "*p*" is true if and only if *p*. But it says nothing about the fact that we aim at making true statements—that we value the truth, in short.

Dummett claims that realism is characterized by the belief that there is something in virtue of which statements are definitely true or false. That is, a realist about a certain type of statement, for example, statements about the distant past, takes them to be either true or false, even if we don't have evidence one way or the other. Thus realists believe that truth is possibly evidence-transcendent, while antirealists deny that truth can transcend what a speaker can verify. In Dummett's view, realism is faced

with a formidable problem. If truth does transcend what we can verify, then we won't necessarily be able to recognize when a statement has that property. And if we can't recognize when a statement has the property of truth, then it is difficult to say whether we actually understand the word "truth" at all. In short, realism makes our understanding of the concept of truth mysterious. And since understanding and truth are closely inter-linked, realism makes understanding, or meaning, mysterious as well.

Dummett suggests that antirealism provides a genuine alternative to the realist picture of the world. But he admits that it has quite radical consequences. Among others, antirealism may force us to reject the law of the excluded middle and the principle of bivalence. It implies that many statements, including statements about the distant past or about the far side of the universe, literally do not have truth values, for they can be neither verified nor falsified. To many philosophers, this seems counter-intuitive. According to our ordinary understanding, statements such as "Caesar stubbed his toe when crossing the Rubicon" do have truth values.<sup>1</sup>

### **Contemporary Pragmatism: Putnam and Rorty**

Hilary Putnam and Richard Rorty are the two most influential philosophers who currently wear the pragmatist label. But while Putnam takes his inspiration from James and is concerned to reconcile pragmatist insights with realism, Rorty looks to Dewey and urges us to leave behind our realist intuitions.

In his essay, Putnam contrasts two overall pictures of reality. Metaphysical realism is the tripartite view that truth is correspondence with reality, that there is a fixed totality of mind-independent objects, and that there is just one true description of the world. The implication is that there is a wide gap between mind and world. Thus, no matter how diligently we may try to believe truly, we can always be wrong. Of course, on this general theoretical picture, there must be some explanation of how thought relates to the world. Yet metaphysical realism, Putnam argues, can't give such an account—that is, a coherent account of truth and reference—that does not end up being decidedly unnatural in its metaphysical commitments.

Putnam's calls his own view internal (and later "pragmatic") realism. Unlike the metaphysical realist, the internal realist believes that the totality of objects is not fixed, because objects themselves exist only relative to conceptual schemes. In Putnam's eyes, this is the consequence of rejecting metaphysical realism. According to Putnam's alternative account of truth, *a proposition is true just when that proposition would be rationally acceptable in ideal epistemic conditions.*

By framing his conception of truth subjunctively and in terms of "ideal" conditions, Putnam overcomes some of the standard objections to other epistemic theories. First, the theory is able to distinguish between propositions that we merely have good reason to believe and propositions that are true. For not all propositions that now pass for true would be rationally accepted were we to have all the relevant facts at hand. Second, Putnam can allow that propositions might have or lack the property of being ideally rationally acceptable even if ideal conditions never actually obtain, since ideal epistemic conditions, like frictionless plains, are just that: ideal. Finally, Putnam's notion of ideal conditions is a fairly broad one: unlike Peirce, he is not claiming that the only truth is what is revealed at the end of scientific inquiry.

While Putnam's position clearly has many strengths that other epistemic theories lack, it still faces an extension of a by-now familiar problem. Some propositions, such as the proposition that the number of stars in the universe at this moment is even, seem incapable of being rationally acceptable or unacceptable even if conditions were ideal. Creatures like us just are not capable of having evidence for such a proposition one way or the other. And yet it seems either true or false. Further, one might worry whether Putnam's account can be informative. For how can we can understand what it means to say that a proposition is rationally justified without at some point invoking a prior concept of truth? For a detailed discussion of these and other important objections to Putnam's position, see Wright (chap. 32) and Alston (chap. 3).

Richard Rorty begins his essay by admitting that pragmatists like himself vacillate between defining truth in terms of justification (ideal or otherwise) and dismissing the problem of truth altogether as a pseudo-problem. This is because the pragmatist can see little practical difference



between truth and justification. When we engage in inquiry, we aim at forming true beliefs, but we can do so only by trying to form justified beliefs. Anything we believe to be true we also believe to be justified, and anything we believe to be justified we also believe to be true. While there may not be much practical difference between truth and justification, Rorty agrees that a distinction can be made. The word “true” has what he calls its “cautionary use.” By distinguishing between justification and truth, we remind ourselves that what is justified now may not be justified later. There is also something of a normative difference between the concepts: unlike justification, truth, Rorty thinks, requires a project of “metaphysical activism” that tries to say how the mind reflects reality.

Rorty opposes metaphysical activism with quietism, a position that he sees in the writings of James and Dewey as well as in the contemporary work of Donald Davidson (chap. 26). Rorty’s quietism takes an essentially deflationist stance towards truth and related concepts like objectivity. Truth, on this account, does no real explanatory work distinct from that done by justification. Thus, Rorty argues against Crispin Wright (chap. 32), we don’t need to think of truth as a distinct aim of inquiry.

After a discussion of Wright’s reason for embracing a metaphysically active position on these issues, Rorty concedes that positions that leave room for a substantive and important concept of truth are bound to feel more intuitive than his own version of pragmatism. Thus Rorty’s position is ultimately normative: in his view, contemporary pragmatists are like atheists in overwhelmingly religious cultures. They must admit that their view is bound to seem strange, since it runs contrary to the mass of public opinion. To make their case, they cannot appeal to that opinion, but rather must rely on pointing out what they take to be its unfortunate consequences. They must argue for their position on pragmatic grounds.

#### Note

1. For discussion of this and other matters concerning Dummett’s antirealism, see Putnam, chap. 30.

#### Further Reading for Part III

Dummett, M. 1978. *Truth and Other Enigmas*. Cambridge: Harvard University Press.

- Dummett, M. 1981. *Frege: Philosophy of Language*. 2nd ed. Cambridge: Harvard University Press.
- Dummett, M. 1983. *The Seas of Language*. Oxford: Oxford University Press.
- Goodman, N. 1978. *Ways of Worldmaking*. Indianapolis: Hackett Publishing Company.
- James, W. 1907 and 1909. *Pragmatism and The Meaning of Truth*. Cambridge: Harvard University Press, 1975.
- Johnston, M. 1993. "Objectivity Refigured: Pragmatism without Verificationism." In *Reality, Representation, and Projection*, edited by John Haldane and Crispin Wright. Oxford: Oxford University Press.
- Peirce, C. 1905. "What Pragmatism Is." *Monist* 15: 161–181.
- Peirce, C. 1932–1933. *Collected Papers*. Vols. 2–4. Cambridge: Harvard University Press.
- Plantinga, A. 1982. "How to Be an Anti-realist." *Proceedings and Addresses of the American Philosophical Association* 56: 47–70.
- Putnam, H. 1981. *Reason, Truth, and History*. Cambridge: Cambridge University Press.
- Putnam, H. 1990. *Realism with a Human Face*. Cambridge: Harvard University Press.
- Putnam, R. A., ed. 1997. *The Cambridge Companion to William James*. Cambridge: Cambridge University Press.
- Rorty, R. 1979. *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.
- Rorty, R. 1993. "Putnam and the Relativist Menace." *Journal of Philosophy* 90, no. 9: 443–461.
- Rorty, R. 1998. *Truth and Progress*. Cambridge: Cambridge University Press.
- Sosa, E. 1993a. "Putnam's Pragmatic Realism." *Journal of Philosophy* 90, no. 12: 605–626.
- Wright, C. 1993. *Realism, Meaning, and Truth*. 2nd ed. Oxford: Basil Blackwell.



---

## How to Make Our Ideas Clear

Charles Sanders Peirce

### I

Whoever has looked into a modern treatise on logic of the common sort, will doubtless remember the two distinctions between *clear* and *obscure* conceptions, and between *distinct* and *confused* conceptions. They have lain in the books now for nigh two centuries, unimproved and unmodified, and are generally reckoned by logicians as among the gems of their doctrine.

A clear idea is defined as one which is so apprehended that it will be recognized wherever it is met with, and so that no other will be mistaken for it. If it fails of this clearness, it is said to be obscure.

This is rather a neat bit of philosophical terminology; yet, since it is clearness that they were defining, I wish the logicians had made their definition a little more plain. Never to fail to recognize an idea, and under no circumstances to mistake another for it, let it come in how recondite a form it may, would indeed imply such prodigious force and clearness of intellect as is seldom met with in this world. On the other hand, merely to have such an acquaintance with the idea as to have become familiar with it, and to have lost all hesitancy in recognizing it in ordinary cases, hardly seems to deserve the name of clearness of apprehension, since after all it only amounts to a subjective feeling of mastery which may be entirely mistaken. I take it, however, that when the logicians speak of “clearness,” they mean nothing more than such a familiarity with an idea, since they regard the quality as but a small merit, which needs to be supplemented by another, which they call *distinctness*.

A distinct idea is defined as one which contains nothing which is not clear. This is technical language; by the *contents* of an idea logicians understand whatever is contained in its definition. So that an idea is *distinctly* apprehended, according to them, when we can give a precise definition of it, in abstract terms. Here the professional logicians leave the subject; and I would not have troubled the reader with what they have to say, if it were not such a striking example of how they have been slumbering through ages of intellectual activity, listlessly disregarding the enginery of modern thought, and never dreaming of applying its lessons to the improvement of logic. It is easy to show that the doctrine that familiar use and abstract distinctness make the perfection of apprehension has its only true place in philosophies which have long been extinct; and it is now time to formulate the method of attaining to a more perfect clearness of thought, such as we see and admire in the thinkers of our own time.

When Descartes set about the reconstruction of philosophy, his first step was to (theoretically) permit skepticism and to discard the practice of the schoolmen of looking to authority as the ultimate source of truth. That done, he sought a more natural fountain of true principles, and professed to find it in the human mind; thus passing, in the directest way, from the method of authority to that of apriority, as described in my first paper.<sup>1</sup> Self-consciousness was to furnish us with our fundamental truths, and to decide what was agreeable to reason. But since, evidently, not all ideas are true, he was led to note, as the first condition of infallibility, that they must be clear. The distinction between an idea *seeming* clear and really being so, never occurred to him. Trusting to introspection, as he did, even a knowledge of external things, why should he question its testimony in respect to the contents of our own minds? But then, I suppose, seeing men, who seemed to be quite clear and positive, holding opposite opinions upon fundamental principles, he was further led to say that clearness of ideas is not sufficient, but that they need also to be distinct, i.e., to have nothing unclear about them. What he probably meant by this (for he did not explain himself with precision) was, that they must sustain the test of dialectical examination; that they must not only seem clear at the outset, but that discussion must never be able to bring to light points of obscurity connected with them.

Such was the distinction of Descartes, and one sees that it was precisely on the level of his philosophy. It was somewhat developed by Leibnitz. This great and singular genius was as remarkable for what he failed to see as for what he saw. That a piece of mechanism could not do work perpetually without being fed with power in some form, was a thing perfectly apparent to him; yet he did not understand that the machinery of the mind can only transform knowledge, but never originate it, unless it be fed with facts of observation. He thus missed the most essential point of the Cartesian philosophy, which is, that to accept propositions which seem perfectly evident to us is a thing which, whether it be logical or illogical, we cannot help doing. Instead of regarding the matter in this way, he sought to reduce the first principles of science to formulas which cannot be denied without self-contradiction, and was apparently unaware of the great difference between his position and that of Descartes. So he reverted to the old formalities of logic, and, above all, abstract definitions played a great part in his philosophy. It was quite natural, therefore, that on observing that the method of Descartes labored under the difficulty that we may seem to ourselves to have clear apprehensions of ideas which in truth are very hazy, no better remedy occurred to him than to require an abstract definition of every important term. Accordingly, in adopting the distinction of *clear* and *distinct* notions, he described the latter quality as the clear apprehension of everything contained in the definition; and the books have ever since copied his words. There is no danger that his chimerical scheme will ever again be overvalued. Nothing new can ever be learned by analyzing definitions. Nevertheless, our existing beliefs can be set in order by this process, and order is an essential element of intellectual economy, as of every other. It may be acknowledged, therefore, that the books are right in making familiarity with a notion the first step toward clearness of apprehension, and the defining of it the second. But in omitting all mention of any higher perspicuity of thought, they simply mirror a philosophy which was exploded a hundred years ago. That much-admired “ornament of logic”—the doctrine of clearness and distinctness—may be pretty enough, but it is high time to relegate to our cabinet of curiosities the antique *bijou*, and to wear about us something better adapted to modern uses.

The very first lesson that we have a right to demand that logic shall teach us is, how to make our ideas clear; and a most important one it is, depreciated only by minds who stand in need of it. To know what we think, to be masters of our own meaning, will make a solid foundation for great and weighty thought. It is most easily learned by those whose ideas are meagre and restricted; and far happier they than such as wallow helplessly in a rich mud of conceptions. A nation, it is true, may, in the course of generations, overcome the disadvantage of an excessive wealth of language and its natural concomitant, a vast, unfathomable deep of ideas. We may see it in history, slowly perfecting its literary forms, sloughing at length its metaphysics, and, by virtue of the untirable patience which is often a compensation, attaining great excellence in every branch of mental acquirement. The page of history is not yet unrolled which is to tell us whether such a people will or will not in the long-run prevail over one whose ideas (like the words of their language) are few, but which possesses a wonderful mastery over those which it has. For an individual, however, there can be no question that a few clear ideas are worth more than many confused ones. A young man would hardly be persuaded to sacrifice the greater part of his thoughts to save the rest; and the muddled head is the least apt to see the necessity of such a sacrifice. Him we can usually only commiserate, as a person with a congenital defect. Time will help him, but intellectual maturity with regard to clearness comes rather late, an unfortunate arrangement of Nature, inasmuch as clearness is of less use to a man settled in life, whose errors have in great measure had their effect, than it would be to one whose path lies before him. It is terrible to see how a single unclear idea, a single formula without meaning, lurking in a young man's head, will sometimes act like an obstruction of inert matter in an artery, hindering the nutrition of the brain, and condemning its victim to pine away in the fullness of his intellectual vigor and in the midst of intellectual plenty. Many a man has cherished for years as his hobby some vague shadow of an idea, too meaningless to be positively false; he has, nevertheless, passionately loved it, has made it his companion by day and by night, and has given to it his strength and his life, leaving all other occupations for its sake, and in short has lived with it and for it, until it has become, as it were, flesh of his flesh and bone of his bone; and then he has waked up some bright morning to find it gone,

clean vanished away like the beautiful Melusina of the fable, and the essence of his life gone with it. I have myself known such a man; and who can tell how many histories of circle-squarers, metaphysicians, astrologers, and what not, may not be told in the old German story?

## II

The principles set forth in the first of these papers lead, at once, to a method of reaching a clearness of thought of a far higher grade than the "distinctness" of the logicians. We have there found that the action of thought is excited by the irritation of doubt, and ceases when belief is attained; so that the production of belief is the sole function of thought. All these words, however, are too strong for my purpose. It is as if I had described the phenomena as they appear under a mental microscope. Doubt and Belief, as the words are commonly employed, relate to religious or other grave discussions. But here I use them to designate the starting of any question, no matter how small or how great, and the resolution of it. If, for instance, in a horse-car, I pull out my purse and find a five-cent nickel and five coppers, I decide, while my hand is going to the purse, in which way I will pay my fare. To call such a question Doubt, and my decision Belief, is certainly to use words very disproportionate to the occasion. To speak of such a doubt as causing an irritation which needs to be appeased, suggests a temper which is uncomfortable to the verge of insanity. Yet, looking at the matter minutely, it must be admitted that, if there is the least hesitation as to whether I shall pay the five coppers or the nickel (as there will be sure to be, unless I act from some previously contracted habit in the matter), though irritation is too strong a word, yet I am excited to such small mental activities as may be necessary to deciding how I shall act. Most frequently doubts arise from some indecision, however momentary, in our action. Sometimes it is not so. I have, for example, to wait in a railway station, and to pass the time I read the advertisements on the walls, I compare the advantages of different trains and different routes which I never expect to take, merely fancying myself to be in a state of hesitancy, because I am bored with having nothing to trouble me. Feigned hesitancy, whether feigned for mere amusement or with a lofty purpose, plays a great part in the pro-



duction of scientific inquiry. However the doubt may originate, it stimulates the mind to an activity which may be slight or energetic, calm or turbulent. Images pass rapidly through consciousness, one incessantly melting into another, until at last, when all is over—it may be in a fraction of a second, in an hour, or after long years—we find ourselves decided as to how we should act under such circumstances as those which occasioned our hesitation. In other words, we have attained belief.

In this process we observe two sorts of elements of consciousness, the distinction between which may best be made clear by means of an illustration. In a piece of music there are the separate notes, and there is the air. A single tone may be prolonged for an hour or a day, and it exists as perfectly in each second of that time as in the whole taken together; so that, as long as it is sounding, it might be present to a sense from which everything in the past was as completely absent as the future itself. But it is different with the air, the performance of which occupies a certain time, during the portions of which only portions of it are played. It consists in an orderliness in the succession of sounds which strike the ear at different times; and to perceive it there must be some continuity of consciousness which makes the events of a lapse of time present to us. We certainly only perceive the air by hearing the separate notes; yet we cannot be said to directly hear it, for we hear only what is present at the instant, and an orderliness of succession cannot exist in an instant. These two sorts of objects, what we are *immediately* conscious of and what we are *mediately* conscious of, are found in all consciousness. Some elements (the sensations) are completely present at every instant so long as they last, while others (like thought) are actions having beginning, middle, and end, and consist in a congruence in the succession of sensations which flow through the mind. They cannot be immediately present to us, but must cover some portion of the past or future. Thought is a thread of melody running through the succession of our sensations.

We may add that just as a piece of music may be written in parts, each part having its own air, so various systems of relationship of succession subsist together between the same sensations. These different systems are distinguished by having different motives, ideas, or functions. Thought is only one such system, for its sole motive, idea, and function, is to produce belief, and whatever does not concern that purpose belongs to some

other system of relations. The action of thinking may incidentally have other results; it may serve to amuse us, for example, and among *dilettanti* it is not rare to find those who have so perverted thought to the purposes of pleasure that it seems to vex them to think that the questions upon which they delight to exercise it may ever get finally settled; and a positive discovery which takes a favorite subject out of the arena of literary debate is met with ill-concealed dislike. This disposition is the very debauchery of thought. But the soul and meaning of thought, abstracted from the other elements which accompany it, though it may be voluntarily thwarted, can never be made to direct itself toward anything but the production of belief. Thought in action has for its only possible motive the attainment of thought at rest; and whatever does not refer to belief is no part of the thought itself.

And what, then, is belief? It is the demi-cadence which closes a musical phrase in the symphony of our intellectual life. We have seen that it has just three properties: First, it is something that we are aware of; second, it appeases the irritation of doubt; and, third, it involves the establishment in our nature of a rule of action, or, say for short, a *habit*. As it appeases the irritation of doubt, which is the motive for thinking, thought relaxes, and comes to rest for a moment when belief is reached. But, since belief is a rule for action, the application of which involves further doubt and further thought, at the same time that it is a stopping-place, it is also a new starting place for thought. That is why I have permitted myself to call it thought at rest, although thought is essentially an action. The *final* upshot of thinking is the exercise of volition, and of this thought no longer forms a part; but belief is only a stadium of mental action, an effect upon our nature due to thought, which will influence future thinking.

The essence of belief is the establishment of a habit, and different beliefs are distinguished by the different modes of action to which they give rise. If beliefs do not differ in this respect, if they appease the same doubt by producing the same rule of action, then no mere differences in the manner of consciousness of them can make them different beliefs, any more than playing a tune in different keys is playing different tunes. Imaginary distinctions are often drawn between beliefs which differ only in their mode of expression;—the wrangling which ensues is real enough, however. To believe that any objects are arranged as in Fig. 1, and to believe that they

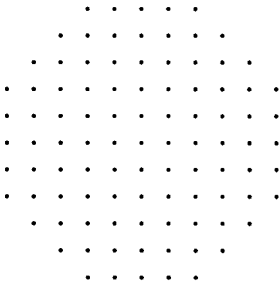


Figure 1

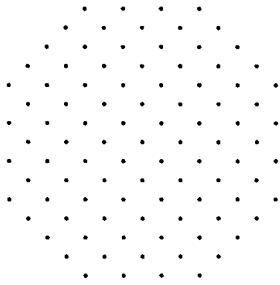


Figure 2

are arranged in Fig. 2, are one and the same belief; yet it is conceivable that a man should assert one proposition and deny the other. Such false distinctions do as much harm as the confusion of beliefs really different, and are among the pitfalls of which we ought constantly to beware, especially when we are upon metaphysical ground. One singular deception of this sort, which often occurs, is to mistake the sensation produced by our own unclearness of thought for a character of the object we are thinking. Instead of perceiving that the obscurity is purely subjective, we fancy that we contemplate a quality of the object which is essentially mysterious; and if our conception be afterward presented to us in a clear form we do not recognize it as the same, owing to the absence of the feeling of unintelligibility. So long as this deception lasts, it obviously puts an impassable barrier in the way of perspicuous thinking; so that it equally interests the opponents of rational thought to perpetuate it, and its adherents to guard against it.

Another such deception is to mistake a mere difference in the grammatical construction of two words for a distinction between the ideas they express. In this pedantic age, when the general mob of writers attend so much more to words than to things, this error is common enough. When I just said that thought is an action, and that it consists in a relation, although a person performs an action but not a relation, which can only be the result of an action, yet there was no inconsistency in what I said, but only a grammatical vagueness.

From all these sophisms we shall be perfectly safe so long as we reflect that the whole function of thought is to produce habits of action; and that whatever there is connected with a thought, but irrelevant to its purpose, is an accretion to it, but no part of it. If there be a unity among our sensations which has no reference to how we shall act on a given occasion, as when we listen to a piece of music, why we do not call that thinking. To develop its meaning, we have, therefore, simply to determine what habits it produces, for what a thing means is simply what habits it involves. Now, the identity of a habit depends on how it might lead us to act, not merely under such circumstances as are likely to arise, but under such as might possibly occur, no matter how improbable they may be. What the habit is depends on when and *how* it causes us to act. As for the *when*, every stimulus to action is derived from perception; as for the *how*, every purpose of action is to produce some sensible result. Thus, we come down to what is tangible and practical, as the root of every real distinction of thought, no matter how subtile it may be; and there is no distinction of meaning so fine as to consist in anything but a possible difference of practice.

To see what this principle leads to, consider in the light of it such a doctrine as that of transubstantiation. The Protestant churches generally hold that the elements of the sacrament are flesh and blood only in a tropical sense; they nourish our souls as meat and the juice of it would our bodies. But the Catholics maintain that they are literally just that; although they possess all the sensible qualities of wafer-cakes and diluted wine. But we can have no conception of wine except what may enter into a belief, either—

1. That this, that, or the other, is wine; or,
2. That wine possesses certain properties.

Such beliefs are nothing but self-notifications that we should, upon occasion, act in regard to such things as we believe to be wine according to the qualities which we believe wine to possess. The occasion of such action would be some sensible perception, the motive of it to produce some sensible result. Thus our action has exclusive reference to what affects the senses, our habit has the same bearing as our action, our belief the same as our habit, our conception the same as our belief; and we can consequently mean nothing by wine but what has certain effects, direct or indirect, upon our senses; and to talk of something as having all the sensible characters of wine, yet being in reality blood, is senseless jargon. Now, it is not my object to pursue the theological question; and having used it as a logical example I drop it, without caring to anticipate the theologian's reply. I only desire to point out how impossible it is that we should have an idea in our minds which relates to anything but conceived sensible effects of things. Our idea of anything is our idea of its sensible effects; and if we fancy that we have any other we deceive ourselves, and mistake a mere sensation accompanying the thought for a part of the thought itself. It is absurd to say that thought has any meaning unrelated to its only function. It is foolish for Catholics and Protestants to fancy themselves in disagreement about the elements of the sacrament, if they agree in regard to all their sensible effects, here or hereafter.

It appears, then, that the rule for attaining the third grade of clearness of apprehension is as follows: Consider what effects, which might conceivably have practical bearings, we conceive the object of our conception to have. Then, our conception of these effects is the whole of our conception of the object.

### III

Let us illustrate this rule by some examples; and, to begin with the simplest one possible, let us ask what we mean by calling a thing *hard*. Evidently that it will not be scratched by many other substances. The whole conception of this quality, as of every other, lies in its conceived effects. There is absolutely no difference between a hard thing and a soft thing so long as they are not brought to the test. Suppose, then, that a diamond could be crystallized in the midst of a cushion of soft cotton,

and should remain there until it was finally burned up. Would it be false to say that that diamond was soft? This seems a foolish question, and would be so, in fact, except in the realm of logic. There such questions are often of the greatest utility as serving to bring logical principles into sharper relief than real discussions ever could. In studying logic we must not put them aside with hasty answers, but must consider them with attentive care, in order to make out the principles involved. We may, in the present ease, modify our question, and ask what prevents us from saying that all hard bodies remain perfectly soft until they are touched, when their hardness increases with the pressure until they are scratched. Reflection will show that the reply is this: there would be no *falsity* in such modes of speech. They would involve a modification of our present usage of speech with regard to the words hard and soft, but not of their meanings. For they represent no fact to be different from what it is; only they involve arrangements of facts which would be exceedingly maladroit. This leads us to remark that the question of what would occur under circumstances which do not actually arise is not a question of fact, but only of the most perspicuous arrangement of them. For example, the question of free-will and fate in its simplest form, stripped of verbiage, is something like this: I have done something of which I am ashamed; could I, by an effort of the will, have resisted the temptation, and done otherwise? The philosophical reply is, that this is not a question of fact, but only of the arrangement of facts. Arranging them so as to exhibit what is particularly pertinent to my question—namely, that I ought to blame myself for having done wrong—it is perfectly true to say that, if I had willed to do otherwise than I did, I should have done otherwise. On the other hand, arranging the facts so as to exhibit another important consideration, it is equally true that, when a temptation has once been allowed to work, it will, if it has a certain force, produce its effect, let me struggle how I may. There is no objection to a contradiction in what would result from a false supposition. The *reductio ad absurdum* consists in showing that contradictory results would follow from a hypothesis which is consequently judged to be false. Many questions are involved in the free-will discussion, and I am far from desiring to say that both sides are equally right. On the contrary, I am of opinion that one side denies important facts, and that the other does not. But what I do say is, that the

above single question was the origin of the whole doubt; that, had it not been for this question, the controversy would never have arisen; and that this question is perfectly solved in the manner which I have indicated. . . .

#### IV

Let us now approach the subject of logic, and consider a conception which particularly concerns it, that of *reality*. Taking clearness in the sense of familiarity, no idea could be clearer than this. Every child uses it with perfect confidence, never dreaming that he does not understand it. As for clearness in its second grade, however, it would probably puzzle most men, even among those of a reflective turn of mind, to give an abstract definition of the real. Yet such a definition may perhaps be reached by considering the points of difference between reality and its opposite, fiction. A figment is a product of somebody's imagination; it has such characters as his thought impresses upon it. That those characters are independent of how you or I think is an external reality. There are, however, phenomena within our own minds, dependent upon our thought, which are at the same time real in the sense that we really think them. But though their characters depend on how we think, they do not depend on what we think those characters to be. Thus, a dream has a real existence as a mental phenomenon, if somebody has really dreamt it; that he dreamt so and so, does not depend on what anybody thinks was dreamt, but is completely independent of all opinion on the subject. On the other hand, considering, not the fact of dreaming, but the thing dreamt, it retains its peculiarities by virtue of no other fact than that it was dreamt to possess them. Thus we may define the real as that whose characters are independent of what anybody may think them to be.

But, however satisfactory such a definition may be found, it would be a great mistake to suppose that it makes the idea of reality perfectly clear. Here, then, let us apply our rules. According to them, reality, like every other quality, consists in the peculiar sensible effects which things partaking of it produce. The only effect which real things have is to cause belief, for all the sensations which they excite emerge into consciousness in the form of beliefs. The question therefore is, how is true belief (or belief in the real) distinguished from false belief (or belief in fiction).

Now, as we have seen in the former paper, the ideas of truth and falsehood, in their full development, appertain exclusively to the scientific method of settling opinion. A person who arbitrarily chooses the propositions which he will adopt can use the word truth only to emphasize the expression of his determination to hold on to his choice. Of course, the method of tenacity never prevailed exclusively; reason is too natural to men for that. But in the literature of the dark ages we find some fine examples of it. When Scotus Erigena is commenting upon a poetical passage in which hellebore is spoken of as having caused the death of Socrates, he does not hesitate to inform the inquiring reader that Helleborus and Socrates were two eminent Greek philosophers, and that the latter having been overcome in argument by the former took the matter to heart and died of it! What sort of an idea of truth could a man have who could adopt and teach, without the qualification of a perhaps, and opinion taken so entirely at random? The real spirit of Socrates, who I hope would have been delighted to have been “overcome in argument,” because he would have learned something by it, is in curious contrast with the naive idea of the glossist, for whom discussion would seem to have been simply a struggle. When philosophy began to awake from its long slumber, and before theology completely dominated it, the practice seems to have been for each professor to seize upon any philosophical position he found unoccupied and which seemed a strong one, to intrench himself in it, and to sally forth from time to time to give battle to the others. Thus, even the scanty records we possess of those disputes enable us to make out a dozen or more opinions held by different teachers at one time concerning the question of nominalism and realism. Read the opening part of the “*Historia Calamitatum*” of Abelard, who was certainly as philosophical as any of his contemporaries, and see the spirit of combat which it breathes. For him, the truth is simply his particular stronghold. When the method of authority prevailed, the truth meant little more than the Catholic faith. All the efforts of the scholastic doctors are directed toward harmonizing their faith in Aristotle and their faith in the Church, and one may search their ponderous folios through without finding an argument which goes any further. It is noticeable that where different faiths flourish side by side, renegades are looked upon with contempt even by the party whose belief they adopt; so completely has the



idea of loyalty replaced that of truth-seeking. Since the time of Descartes, the defect in the conception of truth has been less apparent. Still, it will sometimes strike a scientific man that the philosophers have been less intent on finding out what the facts are, than on inquiring what belief is most in harmony with their system. It is hard to convince a follower of the *a priori* method by adducing facts; but show him that an opinion he is defending is inconsistent with what he has laid down elsewhere, and he will be very apt to retract it. These minds do not seem to believe that disputation is ever to cease; they seem to think that the opinion which is natural for one man is not so for another, and that belief will, consequently, never be settled. In contenting themselves with fixing their own opinions by a method which would lead another man to a different result, they betray their feeble hold of the conception of what truth is.

On the other hand, all the followers of science are fully persuaded that the processes of investigation, if only pushed far enough, will give one certain solution to every question to which they can be applied. One man may investigate the velocity of light by studying the transits of Venus and the aberration of the stars; another by the oppositions of Mars and the eclipses of Jupiter's satellites; a third by the method of Fizeau; a fourth by that of Foucault; a fifth by the motions of the curves of Lissajoux; a sixth, a seventh, an eighth, and a ninth, may follow the different methods of comparing the measures of statical and dynamical electricity. They may at first obtain different results, but, as each perfects his method and his processes, the results will move steadily together toward a destined centre. So with all scientific research. Different minds may set out with the most antagonistic views, but the progress of investigation carries them by a force outside of themselves to one and the same conclusion. This activity of thought by which we are carried, not where we wish, but to a fore-ordained goal, is like the operation of destiny. No modification of the point of view taken, no selection of other facts for study, no natural bent of mind even, can enable a man to escape the predestinate opinion. This great law is embodied in the conception of truth and reality. The opinion which is fated<sup>2</sup> to be ultimately agreed to by all who investigate, is what we mean by the truth, and the object represented in this opinion is the real. That is the way I would explain reality.

But it may be said that this view is directly opposed to the abstract definition which we have given of reality, inasmuch as it makes the characters of the real to depend on what is ultimately thought about them. But the answer to this is that, on the one hand, reality is independent, not necessarily of thought in general, but only of what you or I or any finite number of men may think about it; and that, on the other hand, though the object of the final opinion depends on what that opinion is, yet what that opinion is does not depend on what you or I or any man thinks. Our perversity and that of others may indefinitely postpone the settlement of opinion; it might even conceivably cause an arbitrary proposition to be universally accepted as long as the human race should last. Yet even that would not change the nature of the belief, which alone could be the result of investigation carried sufficiently far; and if, after the extinction of our race, another should arise with faculties and disposition for investigation, that true opinion must be the one which they would ultimately come to. "Truth crushed to earth shall rise again," and the opinion which would finally result from investigation does not depend on how anybody may actually think. But the reality of that which is real does depend on the real fact that investigation is destined to lead, at last, if continued long enough, to a belief in it.

But I may be asked what I have to say to all the minute facts of history, forgotten never to be recovered, to the lost books of the ancients, to the buried secrets.

Full many a gem of purest ray serene

The dark, unfathomed caves of ocean bear;

Full many a flower is born to blush unseen,

And waste its sweetness on the desert air.

Do these things not really exist because they are hopelessly beyond the reach of our knowledge? And then, after the universe is dead (according to the prediction of some scientists), and all life has ceased forever, will not the shock of atoms continue though there will be no mind to know it? To this I reply that, though in no possible state of knowledge can any number be great enough to express the relation between the amount of what rests unknown to the amount of the known, yet it is unphilosophical to suppose that, with regard to any given question (which has

any clear meaning), investigation would not bring forth a solution of it, if it were carried far enough. Who would have said, a few years ago, that we could ever know of what substances stars are made whose light may have been longer in reaching us than the human race has existed? Who can be sure of what we shall not know in a few hundred years? Who can guess what would be the result of continuing the pursuit of science for ten thousand years, with the activity of the last hundred? And if it were to go on for a million, or a billion, or any number of years you please, how is it possible to say that there is any question which might not ultimately be solved?

But it may be objected, "Why make so much of these remote considerations, especially when it is your principle that only practical distinctions have a meaning?" Well, I must confess that it makes very little difference whether we say that a stone on the bottom of the ocean, in complete darkness, is brilliant or not—that is to say, that it *probably* makes no difference, remembering always that that stone *may* be fished up to-morrow. But that there are gems at the bottom of the sea, flowers in the untraveled desert, etc., are propositions which, like that about a diamond being hard when it is not pressed, concern much more the arrangement of our language than they do the meaning of our ideas.

It seems to me, however, that we have, by the application of our rule, reached so clear an apprehension of what we mean by reality, and of the fact which the idea rests on, that we should not, perhaps, be making a pretension so presumptuous as it would be singular, if we were to offer a metaphysical theory of existence for universal acceptance among those who employ the scientific method of fixing belief. However, as metaphysics is a subject much more curious than useful, the knowledge of which, like that of a sunken reef, serves chiefly to enable us to keep clear of it, I will not trouble the reader with any more Ontology at this moment. I have already been led much further into that path than I should have desired; and I have given the reader such a dose of mathematics, psychology, and all that is most abstruse, that I fear he may already have left me, and that what I am now writing is for the compositor and proof-reader exclusively. I trusted to the importance of the subject. There is no royal road to logic, and really valuable ideas can only be had at the price of close attention. But I know that in the matter of ideas

the public prefer the cheap and nasty; and in my next paper I am going to return to the easily intelligible, and not wander from it again. The reader who has been at the pains of wading through this month's paper, shall be rewarded in the next one by seeing how beautifully what has been developed in this tedious way can be applied to the ascertainment of the rules of scientific reasoning.

We have, hitherto, not crossed the threshold of scientific logic. It is certainly important to know how to make our ideas clear, but they may be ever so clear without being true. How to make them so, we have next to study. How to give birth to those vital and procreative ideas which multiply into a thousand forms and diffuse themselves everywhere, advancing civilization and making the dignity of man, is an art not yet reduced to rules, but of the secret of which the history of science affords some hints.

#### Notes

1. Peirce published "The Fixation of Belief" in *Popular Science Monthly* in 1877, a year before "How to Make Our Ideas Clear." See the *Collected Papers of Charles Sanders Peirce*, vol. 5, ed. Charles Hartshorne and Paul Weiss (Cambridge: Harvard University Press, 1965), pp. 223–47.—Ed.
2. Fate means merely that which is sure to come true, and can nohow be avoided. It is a superstition to suppose that a certain sort of events are ever fated, and it is another to suppose that the word fate can never be freed from its superstitious taint. We are all fated to die.



---

## Pragmatism's Conception of Truth

William James

When Clerk-Maxwell was a child it is written that he had a mania for having everything explained to him, and that when people put him off with vague verbal accounts of any phenomenon he would interrupt them impatiently by saying, 'Yes; but I want you to tell me the *particular* go of it!' Had his question been about truth, only a pragmatist could have told him the particular go of it. I believe that our contemporary pragmatists, especially Messrs. Schiller and Dewey, have given the only tenable account of this subject. It is a very ticklish subject, sending subtle rootlets into all kinds of crannies, and hard to treat in the sketchy way that alone befits a public lecture. But the Schiller-Dewey view of truth has been so ferociously attacked by rationalistic philosophers, and so abominably misunderstood, that here, if anywhere, is the point where a clear and simple statement should be made.

I fully expect to see the pragmatist view of truth run through the classic stages of a theory's career. First, you know, a new theory is attacked as absurd; then it is admitted to be true, but obvious and insignificant; finally it is seen to be so important that its adversaries claim that they themselves discovered it. Our doctrine of truth is at present in the first of these three stages, with symptoms of the second stage having begun in certain quarters. I wish that this lecture might help it beyond the first stage in the eyes of many of you.

Truth, as any dictionary will tell you, is a property of certain of our ideas. It means their 'agreement', as falsity means their disagreement, with 'reality'. Pragmatists and intellectualists both accept this definition as a matter of course. They begin to quarrel only after the question is raised as to what may precisely be meant by the term 'agreement', and

what by the term 'reality', when reality is taken as something for our ideas to agree with.

In answering these questions the pragmatists are more analytic and painstaking, the intellectualists more offhand and irreflective. The popular notion is that a true idea must copy its reality. Like other popular views, this one follows the analogy of the most usual experience. Our true ideas of sensible things do indeed copy them. Shut your eyes and think of yonder clock on the wall, and you get just such a true picture or copy of its dial. But your idea of its 'works' (unless you are a clockmaker) is much less of a copy, yet it passes muster, for it in no way clashes with the reality. Even though it should shrink to the mere word 'works', that word still serves you truly; and when you speak of the 'time-keeping function' of the clock, or of its spring's 'elasticity', it is hard to see exactly what your ideas can copy.

You perceive that there is a problem here. Where our ideas cannot copy definitely their object, what does agreement with that object mean? Some idealists seem to say that they are true whenever they are what God means that we ought to think about that object. Others hold the copy-view all through, and speak as if our ideas possessed truth just in proportion as they approach to being copies of the Absolute's eternal way of thinking.

These views, you see, invite pragmatistic discussion. But the great assumption of the intellectualists is that truth means essentially an inert static relation. When you've got your true idea of anything, there's an end of the matter. You're in possession; you *know*; you have fulfilled your thinking destiny. You are where you ought to be mentally; you have obeyed your categorical imperative; and nothing more need follow on that climax of your rational destiny. Epistemologically you are in stable equilibrium.

Pragmatism, on the other hand, asks its usual question. 'Grant an idea or belief to be true,' it says, 'what concrete difference will its being true make in any one's actual life? How will the truth be realized? What experiences will be different from those which would obtain if the belief were false? What, in short, is the truth's cash-value in experiential terms?'

The moment pragmatism asks this question, it sees the answer: *True ideas are those that we can assimilate, validate, corroborate and verify.*

*False ideas are those that we can not.* That is the practical difference it makes to us to have true ideas; that, therefore, is the meaning of truth, for it is all that truth is known-as.

This thesis is what I have to defend. The truth of an idea is not a stagnant property inherent in it. Truth *happens* to an idea. It *becomes* true, is *made* true by events. Its verity *is* in fact an event, a process: the process namely of its verifying itself, its veri-*fication*. Its validity is the process of its valid-*ation*.

But what do the words verification and validation themselves pragmatically mean? They again signify certain practical consequences of the verified and validated idea. It is hard to find any one phrase that characterizes these consequences better than the ordinary agreement-formula—just such consequences being what we have in mind whenever we say that our ideas ‘agree’ with reality. They lead us, namely, through the acts and other ideas which they instigate, into or up to, or towards, other parts of experience with which we feel all the while—such feeling being among our potentialities—that the original ideas remain in agreement. The connexions and transitions come to us from point to point as being progressive, harmonious, satisfactory. This function of agreeable leading is what we mean by an idea’s verification. Such an account is vague and it sounds at first quite trivial, but it has results which it will take the rest of my hour to explain.

Let me begin by reminding you of the fact that the possession of true thoughts means everywhere the possession of invaluable instruments of action; and that our duty to gain truth, so far from being a blank command from out of the blue, or a ‘stunt’ self-imposed by our intellect, can account for itself by excellent practical reasons.

The importance to human life of having true beliefs about matters of fact is a thing too notorious. We live in a world of realities that can be infinitely useful or infinitely harmful. Ideas that tell us which of them to expect count as the true ideas in all this primary sphere of verification, and the pursuit of such ideas is a primary human duty. The possession of truth, so far from being here an end in itself, is only a preliminary means towards other vital satisfactions. If I am lost in the woods and starved, and find what looks like a cow-path, it is of the utmost importance that I



should think of a human habitation at the end of it, for if I do so and follow it, I save myself. The true thought is useful here because the house which is its object is useful. The practical value of true ideas is thus primarily derived from the practical importance of their objects to us. Their objects are, indeed, not important at all times. I may on another occasion have no use for the house; and then my idea of it, however verifiable, will be practically irrelevant, and had better remain latent. Yet since almost any object may some day become temporarily important, the advantage of having a general stock of *extra* truths, of ideas that shall be true of merely possible situations, is obvious. We store such extra truths away in our memories, and with the overflow we fill our books of reference. Whenever such an extra truth becomes practically relevant to one of our emergencies, it passes from cold-storage to do work in the world and our belief in it grows active. You can say of it then either that 'it is useful because it is true' or that 'it is true because it is useful'. Both these phrases mean exactly the same thing, namely that here is an idea that gets fulfilled and can be verified. True is the name for whatever idea starts the verification-process, useful is the name for its completed function in experience. True ideas would never have been singled out as such, would never have acquired a class-name, least of all a name suggesting value, unless they had been useful from the outset in this way.

From this simple cue pragmatism gets her general notion of truth as something essentially bound up with the way in which one moment in our experience may lead us towards other moments which it will be worth while to have been led to. Primarily, and on the common-sense level, the truth of a state of mind means this function of a *leading that is worth while*. When a moment in our experience, of any kind whatever, inspires us with a thought that is true, that means that sooner or later we dip by that thought's guidance into the particulars of experience again and make advantageous connexion with them. This is a vague enough statement, but I beg you to retain it, for it is essential.

Our experience meanwhile is all shot through with regularities. One bit of it can warn us to get ready for another bit, can 'intend' or be 'significant of' that remoter object. The object's advent is the significance's verification. Truth, in these cases, meaning nothing but eventual verification,

is manifestly incompatible with waywardness on our part. Woe to him whose beliefs play fast and loose with the order which realities follow in his experience; they will lead him nowhere or else make false connexions.

By 'realities' or 'objects' here, we mean either things of common sense, sensibly present, or else common-sense relations, such as dates, places, distances, kinds, activities. Following our mental image of a house along the cow-path, we actually come to see the house; we get the image's full verification. *Such simply and fully verified leadings are certainly the originals and prototypes of the truth-process.* Experience offers indeed other forms of truth-process, but they are all conceivable as being primary verifications arrested, multiplied or substituted one for another.

Take, for instance, yonder object on the wall. You and I consider it to be a 'clock', although no one of us has seen the hidden works that make it one. We let our notion pass for true without attempting to verify. If truths mean verification-process essentially, ought we then to call such unverified truths as this abortive? No, for they form the overwhelmingly large number of the truths we live by. Indirect as well as direct verifications pass muster. Where circumstantial evidence is sufficient, we can go without eyewitnessing. Just as we here assume Japan to exist without ever having been there, because it *works* to do so, everything we know conspiring with the belief, and nothing interfering, so we assume that thing to be a clock. We *use* it as a clock, regulating the length of our lecture by it. The verification of the assumption here means its leading to no frustration or contradiction. *Verifiability* of wheels and weights and pendulum is as good as verification. For one truth-process completed there are a million in our lives that function in this state of nascency. They turn us *towards* direct verification; lead us into the *surroundings* of the objects they envisage; and then, if everything runs on harmoniously, we are so sure that verification is possible that we omit it, and are usually justified by all that happens.

Truth lives, in fact, for the most part on a credit system. Our thoughts and beliefs 'pass', so long as nothing challenges them, just as banknotes pass so long as nobody refuses them. But this all points to direct face-to-face verifications somewhere, without which the fabric of truth collapses like a financial system with no cash-basis whatever. You accept my verification of one thing, I yours of another. We trade on each other's truth.

But beliefs verified concretely by *somebody* are the posts of the whole superstructure.

Another great reason—beside economy of time—for waiving complete verification in the usual business of life is that all things exist in kinds and not singly. Our world is found once for all to have that peculiarity. So that when we have once directly verified our ideas about one specimen of a kind, we consider ourselves free to apply them to other specimens without verification. A mind that habitually discerns the kind of thing before it, and acts by the law of the kind immediately, without pausing to verify, will be a ‘true’ mind in ninety-nine out of a hundred emergencies, proved so by its conduct fitting everything it meets, and getting no refutation.

*Indirectly or only potentially verifying processes may thus be true as well as full verification-processes.* They work as true processes would work, give us the same advantages, and claim our recognition for the same reasons. All this on the common-sense level of matters of fact, which we are alone considering.

But matters of fact are not our only stock in trade. *Relations among purely mental ideas* form another sphere where true and false beliefs obtain, and here the beliefs are absolute, or unconditional. When they are true they bear the name either of definitions or of principles. It is either a principle or a definition that 1 and 1 make 2, that 2 and 1 make 3, and so on; that white differs less from grey than it does from black; that when the cause begins to act the effect also commences. Such propositions hold of all possible ‘ones’, of all conceivable ‘whites’ and ‘greys’ and ‘causes’. The objects here are mental objects. Their relations are perceptually obvious at a glance, and no sense-verification is necessary. Moreover, once true, always true, of those same mental objects. Truth here has an ‘eternal’ character. If you can find a concrete thing anywhere that is ‘one’ or ‘white’ or ‘grey’ or an ‘effect,’ then your principles will everlastingly apply to it. It is but a case of ascertaining the kind, and then applying the law of its kind to the particular object. You are sure to get truth if you can but name the kind rightly, for your mental relations hold good of everything of that kind without exception. If you then, never-

theless, failed to get truth concretely, you would say that you had classed your real objects wrongly.

In this realm of mental relations, truth again is an affair of leading. We relate one abstract idea with another, framing in the end great systems of logical and mathematical truth, under the respective terms of which the sensible facts of experience eventually arrange themselves, so that our eternal truths hold good of realities also. This marriage of fact and theory is endlessly fertile. What we say is here already true in advance of special verification, *if we have subsumed our objects rightly*. Our ready-made ideal framework for all sorts of possible objects follows from the very structure of our thinking. We can no more play fast and loose with these abstract relations than we can do so with our sense-experiences. They coerce us; we must treat them consistently, whether or not we like the results. The rules of addition apply to our debts as rigorously as to our assets. The hundredth decimal of  $\pi$ , the ratio of the circumference to its diameter, is predetermined ideally now, though no one may have computed it. If we should ever need the figure in our dealings with an actual circle we should need to have it given rightly, calculated by the usual rules; for it is the same kind of truth that those rules elsewhere calculate.

Between the coercions of the sensible order and those of the ideal order, our mind is thus wedged tightly. Our ideas must agree with realities, be such realities concrete or abstract, be they facts or be they principles, under penalty of endless inconsistency and frustration.

So far, intellectualists can raise no protest. They can only say that we have barely touched the skin of the matter.

Realities mean, then, either concrete facts, or abstract kinds of thing and relations perceived intuitively between them. They furthermore and thirdly mean, as things that new ideas of ours must no less take account of, the whole body of other truths already in our possession. But what now does 'agreement' with such threefold realities mean?—to use again the definition that is current.

Here it is that pragmatism and intellectualism begin to part company. Primarily, no doubt, to agree means to copy, but we saw that the mere word 'clock' would do instead of a mental picture of its works, and that

of many realities our ideas can only be symbols and not copies. 'Past time', 'power', 'spontaneity'—how can our mind copy such realities?

To 'agree' in the widest sense with a reality *can only mean to be guided either straight up to it or into its surroundings, or to be put into such working touch with it as to handle either it or something connected with it better than if we disagreed*. Better either intellectually or practically! And often agreement will only mean the negative fact that nothing contradictory from the quarter of that reality comes to interfere with the way in which our ideas guide us elsewhere. To copy a reality is, indeed, one very important way of agreeing with it, but it is far from being essential. The essential thing is the process of being guided. Any idea that helps us to *deal*, whether practically or intellectually, with either the reality or its belongings, that doesn't entangle our progress in frustrations, that *fits*, in fact, and adapts our life to the reality's whole setting, will agree sufficiently to meet the requirement. It will hold true of that reality.

Thus, *names* are just as 'true' or 'false' as definite mental pictures are. They set up similar verification-processes, and lead to fully equivalent practical results.

All human thinking gets discursified; we exchange ideas; we lend and borrow verifications, get them from one another by means of social intercourse. All truth thus gets verbally built out, stored up, and made available for every one. Hence, we must *talk* consistently just as we must *think* consistently: for both in talk and thought we deal with kinds. Names are arbitrary, but once understood they must be kept to. We mustn't now call Abel 'Cain' or Cain 'Abel'. If we do, we ungear ourselves from the whole book of Genesis, and from all its connexions with the universe of speech and fact down to the present time. We throw ourselves out of whatever truth that entire system of speech and fact may embody.

The overwhelming majority of our true ideas admit of no direct or face-to-face verification—those of past history, for example, as of Cain and Abel. The stream of time can be remounted only verbally, or verified indirectly by the present prolongations or effects of what the past harbored. Yet if they agree with these verbalities and effects, we can know that our ideas of the past are true. *As true as past time itself was*, so true was Julius Cæsar, so true were antediluvian monsters, all in their proper

dates and settings. That past time itself was, is guaranteed by its coherence with everything that's present. True as the present *is*, the past *was* also.

Agreement thus turns out to be essentially an affair of leading—leading that is useful because it is into quarters that contain objects that are important. True ideas lead us into useful verbal and conceptual quarters as well as directly up to useful sensible termini. They lead to consistency, stability and flowing human intercourse. They lead away from eccentricity and isolation, from foiled and barren thinking. The untrammelled flowing of the leading-process, its general freedom from clash and contradiction, passes for its indirect verification; but all roads lead to Rome, and in the end and eventually, all true processes must lead to the face of directly verifying sensible experiences *somewhere*, which somebody's ideas have copied.

Such is the large loose way in which the pragmatist interprets the word agreement. He treats it altogether practically. He lets it cover any process of conduction from a present idea to a future terminus, provided only it run prosperously. It is only thus that 'scientific' ideas, flying as they do beyond common sense, can be said to agree with their realities. It is, as I have already said, *as if* reality were made of ether, atoms or electrons, but we mustn't think so literally. The term 'energy' doesn't even pretend to stand for anything 'objective'. It is only a way of measuring the surface of phenomena so as to string their changes on a simple formula.

Yet in the choice of these man-made formulas we can not be capricious with impunity any more than we can be capricious on the common-sense practical level. We must find a theory that will *work*; and that means something extremely difficult; for our theory must mediate between all previous truths and certain new experiences. It must derange common sense and previous belief as little as possible, and it must lead to some sensible terminus or other that can be verified exactly. To 'work' means both these things; and the squeeze is so tight that there is little loose play for any hypothesis. Our theories are wedged and controlled as nothing else is. Yet sometimes alternative theoretic formulas are equally compatible with all the truths we know, and then we choose between them for subjective reasons. We choose the kind of theory to which we are already partial; we follow 'elegance' or 'economy'. Clerk-Maxwell somewhere

says it would be 'poor scientific taste' to choose the more complicated of two equally well-evidenced conceptions; and you will all agree with him. Truth in science is what gives us the maximum possible sum of satisfactions, taste included, but consistency both with previous truth and with novel fact is always the most imperious claimant.

I have led you through a very sandy desert. But now, if I may be allowed so vulgar an expression, we begin to taste the milk in the coconut. Our rationalist critics here discharge their batteries upon us, and to reply to them will take us out from all this dryness into full sight of a momentous philosophical alternative.

Our account of truth is an account of truths in the plural, of processes of leading, realized *in rebus*, and having only this quality in common, that they *pay*. They pay by guiding us into or towards some part of a system that dips at numerous points into sense-percepts, which we may copy mentally or not, but with which at any rate we are now in the kind of commerce vaguely designated as verification. Truth for us is simply a collective name for verification-processes, just as health, wealth, strength, etc., are names for other processes connected with life, and also pursued because it pays to pursue them. Truth is *made*, just as health, wealth, and strength are made, in the course of experience.

Here rationalism is instantaneously up in arms against us. I can imagine a rationalist to talk as follows:

'Truth is not made,' he will say; 'it absolutely obtains, being a unique relation that does not wait upon any process, but shoots straight over the head of experience, and hits its reality every time. Our belief that yon thing on the wall is a clock is true already, although no one in the whole history of the world should verify it. The bare quality of standing in that transcendent relation is what makes any thought true that possesses it, whether or not there be verification. You pragmatists put the cart before the horse in making truth's being reside in verification-processes. These are merely signs of its being, merely our lame ways of ascertaining after the fact, which of our ideas already has possessed the wondrous quality. The quality itself is timeless, like all essences and natures. Thoughts partake of it directly, as they partake of falsity or of irrelevancy. It can't be analysed away into pragmatic consequences.'

The whole plausibility of this rationalist tirade is due to the fact to which we have already paid so much attention. In our world, namely, abounding as it does in things of similar kinds and similarly associated, one verification serves for others of its kind, and one great use of knowing things is to be led not so much to them as to their associates, especially to human talk about them. The quality of truth, obtaining *ante rem*, pragmatically means, then, the fact that in such a world innumerable ideas work better by their indirect or possible than by their direct and actual verification. Truth *ante rem* means only verifiability, then; or else it is a case of the stock rationalist trick of treating the *name* of a concrete phenomenal reality as an independent prior entity, and placing it behind the reality as its explanation. Professor Mach quotes somewhere an epigram of Lessing's:

Sagt Hänschen Schlau zu Vetter Fritz,  
 'Wie kommt es, Vetter Fritzen,  
 Dass grad' die Reichsten in der Welt,  
 Das meiste Geld besitzen?'

Hänschen Schlau here treats the principle 'wealth' as something distinct from the facts denoted by the man's being rich. It antedates them; the facts become only a sort of secondary coincidence with the rich man's essential nature.

In the case of 'wealth' we all see the fallacy. We know that wealth is but a name for concrete processes that certain men's lives play a part in, and not a natural excellence found in Messrs. Rockefeller and Carnegie, but not in the rest of us.

Like wealth, health also lives *in rebus*. It is a name for processes, and digestion, circulation, sleep, etc., that go on happily, though in this instance we are more inclined to think of it as a principle and to say the man digests and sleeps so well *because* he is so healthy.

With 'strength' we are, I think, more rationalistic still, and decidedly inclined to treat it as an excellence pre-existing in the man and explanatory of the herculean performances of his muscles.

With 'truth' most people go over the border entirely, and treat the rationalistic account as self-evident. But really all these words in *th* are exactly similar. Truth exists *ante rem* just as much and as little as the other things do.



The scholastics, following Aristotle, made much of the distinction between habit and act. Health *in actu* means, among other things, good sleeping and digesting. But a healthy man need not always be sleeping, or always digesting, any more than a wealthy man need be always handing money, or a strong man always lifting weights. All such qualities sink to the status of 'habits' between their times of exercise; and similarly truth becomes a habit of certain of our ideas and beliefs in their intervals of rest from their verifying activities. But those activities are the root of the whole matter, and the condition of there being any habit to exist in the intervals.

*'The true', to put it very briefly, is only the expedient in the way of our thinking, just as 'the right' is only the expedient in the way of our behaving.* Expedient in almost any fashion; and expedient in the long run and on the whole of course; for what meets expediently all the experience in sight won't necessarily meet all farther experiences equally satisfactorily. Experience, as we know, has ways of *boiling over*, and making us correct our present formulas.

The 'absolutely' true, meaning what no further experience will ever alter, is that ideal vanishing-point towards which we imagine that all our temporary truths will some day converge. It runs on all fours with the perfectly wise man, and with the absolutely complete experience; and, if these ideals are ever realized, they will all be realized together. Meanwhile we have to live today by what truth we can get today, and be ready tomorrow to call it falsehood. Ptolemaic astronomy, Euclidean space, Aristotelian logic, scholastic metaphysics, were expedient for centuries, but human experience has boiled over those limits, and we now call these things only relatively true, or true within those borders of experience. 'Absolutely' they are false; for we know that those limits were casual, and might have been transcended by past theorists just as they are by present thinkers.

When new experiences lead to retrospective judgements, using the past tense, what these judgements utter *was* true, even though no past thinker had been led there. We live forwards, a Danish thinker has said, but we understand backwards. The present sheds a backward light on the world's previous processes. They may have been truth-processes for the actors in them. They are not so for one who knows the later revelations of the story.

This regulative notion of a potential better truth to be established later, possibly to be established some day absolutely, and having powers of retroactive legislation, turns its face, like all pragmatist notions, towards concreteness of fact, and towards the future. Like the half-truths, the absolute truth will have to be *made*, made as a relation incidental to the growth of a mass of verification-experience, to which the half-true ideas are all along contributing their quota.

I have already insisted on the fact that truth is made largely out of previous truths. Men's beliefs at any time are so much experience *funded*. But the beliefs are themselves parts of the sum total of the world's experience, and become matter, therefore, for the next day's funding operations. So far as reality means experienceable reality, both it and the truths men gain about it are everlastingly in process of mutation—mutation towards a definite goal, it may be—but still mutation.

Mathematicians can solve problems with two variables. On the Newtonian theory, for instance, acceleration varies with distance, but distance also varies with acceleration. In the realm of truth-processes facts come independently and determine our beliefs provisionally. But these beliefs make us act, and as fast as they do so, they bring into sight or into existence new facts which redetermine the beliefs accordingly. So the whole coil and ball of truth, as it rolls up, is the product of a double influence. Truths emerge from facts; but they dip forward into facts again and add to them; which facts again create or reveal new truth (the word is indifferent) and so on indefinitely. The 'facts' themselves meanwhile are not *true*. They simply *are*. Truth is the function of the beliefs that start and terminate among them.

The case is like a snowball's growth, due as it is to the distribution of the snow on the one hand, and to the successive pushes of the boys on the other, with these factors co-determining each other incessantly.

The most fateful point of difference between being a rationalist and being a pragmatist is now fully in sight. Experience is in mutation, and our psychological ascertainments of truth are in mutation—so much rationalism will allow; but never that either reality itself or truth itself is mutable. Reality stands complete and ready-made from all eternity, rationalism insists, and the agreement of our ideas with it is that unique unanalyzable virtue in them of which she has already told us. As that

intrinsic excellence, their truth has nothing to do with our experiences. It adds nothing to the content of experience. It makes no difference to reality itself; it is supervenient, inert, static, a reflexion merely. It doesn't *exist*, it *holds* or *obtains*, it belongs to another dimension from that of either facts or fact-relations, belongs, in short, to the epistemological dimension—and with that big word rationalism closes the discussion.

Thus, just as pragmatism faces forward to the future, so does rationalism here again face backward to a past eternity. True to her inveterate habit, rationalism reverts to 'principles', and thinks that when an abstraction once is named, we own an oracular solution.

The tremendous pregnancy in the way of consequences for life of this radical difference of outlook will only become apparent in my later lectures. I wish meanwhile to close this lecture by showing that rationalism's sublimity does not save it from inanity.

When, namely, you ask rationalists, instead of accusing pragmatism of desecrating the notion of truth, to define it themselves by saying exactly what *they* understand by it, the only positive attempts I can think of are these two:

1. 'Truth is the system of propositions which have an unconditional claim to be recognized as valid.'<sup>1</sup>
2. Truth is a name for all those judgements which we find ourselves under obligation to make by a kind of imperative duty.<sup>2</sup>

The first thing that strikes one in such definitions is their unutterable triviality. They are absolutely true, of course, but absolutely insignificant until you handle them pragmatically. What do you mean by 'claim' here, and what do you mean by 'duty'? As summary names for the concrete reasons why thinking in true ways is overwhelmingly expedient and good for mortal men, it is all right to talk of claims on reality's part to be agreed with, and of obligations on our part to agree. We feel both the claims and the obligations, and we feel them for just those reasons.

But the rationalists who talk of claim and obligation *expressly say that they have nothing to do with our practical interests or personal reasons*. Our reasons for agreeing are psychological facts, they say, relative to each thinker, and to the accidents of his life. They are his evidence merely, they are no part of the life of truth itself. That life transacts itself

in a purely logical or epistemological, as distinguished from a psychological, dimension, and its claims antedate and exceed all personal motivations whatsoever. Though neither man nor God should ever ascertain truth, the word would still have to be defined as that which *ought* to be ascertained and recognized.

There never was a more exquisite example of an idea abstracted from the concretes of experience and then used to oppose and negate what it was abstracted from.

Philosophy and common life abound in similar instances. The 'sentimentalist fallacy' is to shed tears over abstract justice and generosity, beauty, etc., and never to know these qualities when you meet them in the street, because the circumstances make them vulgar. Thus I read in the privately printed biography of an eminently rationalistic mind: 'It was strange that with such admiration for beauty in the abstract, my brother had no enthusiasm for fine architecture, for beautiful painting, or for flowers.' And in almost the last philosophic work I have read, I find such passages as the following: 'Justice is ideal, solely ideal. Reason conceives that it ought to exist, but experience shows that it can not. . . . Truth, which ought to be, can not be. . . . Reason is deformed by experience. As soon as reason enters experience it becomes contrary to reason.'

The rationalist's fallacy here is exactly like the sentimentalist's. Both extract a quality from the muddy particulars of experience, and find it so pure when extracted that they contrast it with each and all its muddy instances as an opposite and higher nature. All the while it is *their* nature. It is the nature of truths to be validated, verified. It pays for our ideas to be validated. Our obligation to seek truth is part of our general obligation to do what pays. The payments true ideas bring are the sole why of our duty to follow them. Identical whys exist in the case of wealth and health.

Truth makes no other kind of claim and imposes no other kind of ought than health and wealth do. All these claims are conditional; the concrete benefits we gain are what we mean by calling the pursuit a duty. In the case of truth, untrue beliefs work as perniciously in the long run as true beliefs work beneficially. Talking abstractly, the quality 'true' may thus be said to grow absolutely precious and the quality 'untrue' abso-

lutely damnable: the one may be called good, the other bad, unconditionally. We ought to think the true, we ought to shun the false, imperatively.

But if we treat all this abstraction literally and oppose it to its mother soil in experience, see what a preposterous position we work ourselves into.

We cannot then take a step forward in our actual thinking. When shall I acknowledge this truth and when that? Shall the acknowledgment be loud?—or silent? If sometimes loud, sometimes silent, which *now*? When may a truth go into cold-storage in the encyclopedia? and when shall it come out for battle? Must I constantly be repeating the truth ‘twice two are four’ because of its eternal claim on recognition? or is it sometimes irrelevant? Must my thoughts dwell night and day on my personal sins and blemishes, because I truly have them?—or may I sink and ignore them in order to be a decent social unit, and not a mass of morbid melancholy and apology?

It is quite evident that our obligation to acknowledge truth, so far from being unconditional, is tremendously conditioned. Truth with a big T, and in the singular, claims abstractly to be recognized, of course; but concrete truths in the plural need be recognized only when their recognition is expedient. A truth must always be preferred to a falsehood when both relate to the situation; but when neither does, truth is as little of a duty as falsehood. If you ask me what o’clock it is and I tell you that I live at 95 Irving Street, my answer may indeed be true, but you don’t see why it is my duty to give it. A false address would be as much to the purpose.

With this admission that there are conditions that limit the application of the abstract imperative, *the pragmatistic treatment of truth sweeps back upon us in its fulness*. Our duty to agree with reality is seen to be grounded in a perfect jungle of concrete expediencies.

When Berkeley had explained what people meant by matter, people thought that he denied matter’s existence. When Messrs. Schiller and Dewey now explain what people mean by truth, they are accused of denying *its* existence. These pragmatists destroy all objective standards, critics say, and put foolishness and wisdom on one level. A favourite formula for describing Mr Schiller’s doctrines and mine is that we are persons who think that by saying whatever you find it pleasant to say and calling it truth you fulfil every pragmatistic requirement.

I leave it to you to judge whether this be not an impudent slander. Pent in, as the pragmatist more than anyone else sees himself to be, between the whole body of funded truths squeezed from the past and the coercions of the world of sense about him, who so well as he feels the immense pressure of objective control under which our minds perform their operations? If anyone imagines that this law is lax, let him keep its commandment one day, says Emerson. We have heard much of late of the uses of the imagination in science. It is high time to urge the use of a little imagination in philosophy. The unwillingness of some of our critics to read any but the silliest of possible meanings into our statements is as discreditable to their imaginations as anything I know in recent philosophic history. Schiller says the true is that which 'works'. Thereupon he is treated as one who limits verification to the lowest material utilities. Dewey says truth is what gives 'satisfaction'. He is treated as one who believes in calling everything true which, if it were true, would be pleasant.

Our critics certainly need more imagination of realities. I have honestly tried to stretch my own imagination and to read the best possible meaning into the rationalist conception, but I have to confess that it still completely baffles me. The notion of a reality calling on us to 'agree' with it, and that for no reasons, but simply because its claim is 'unconditional' or 'transcendent', is one that I can make neither head nor tail of. I try to imagine myself as the sole reality in the world, and then to imagine what more I would 'claim' if I were allowed to. If you suggest the possibility of my claiming that a mind should come into being from out of the void inane and stand and *copy* me, I can indeed imagine what the copying might mean, but I can conjure up no motive. What good it would do me to be copied, or what good it would do that mind to copy me, if further consequences are expressly and in principle ruled out as motives for the claim (as they are by our rationalist authorities) I can not fathom. When the Irishman's admirers ran him along to the place of banquet in a sedan chair with no bottom, he said, 'Faith, if it wasn't for the honour of the thing, I might as well have come on foot.' So here: but for the honour of the thing, I might as well have remained uncopied. Copying is one genuine mode of knowing (which for some strange reason our contemporary transcendentalists seem to be tumbling over each other to repudiate); but when we get beyond copying, and fall back on unnamed

forms of agreeing that are expressly denied to be either copyings or leadings or fittings, or any other processes pragmatically definable, the *what* of the 'agreement' claimed becomes as unintelligible as the *why* of it. Neither content nor motive can be imagined for it. It is an absolutely meaningless abstraction.<sup>3</sup>

Surely in this field of truth it is the pragmatists and not the rationalists who are the more genuine defenders of the universe's rationality.

#### Notes

1. A. E. Taylor, *Philosophical Review*, 14: 288.
2. H. Rickert, *Der Gegenstand der Erkenntniss: Einfuhrung in die tranzendentale philosophie* (Tübingen, 1904), ch. on 'Die Urtheilsnothwendigkeit'.
3. I am not forgetting that Professor Rickert long ago gave up the whole notion of truth being founded on agreement with reality. Reality according to him, is whatever agrees with truth, and truth is founded solely on our primal duty. This fantastic flight, together with Mr Joachim's candid confession of failure in his book *The Nature of Truth* (Oxford, 1906), seems to me to mark the bankruptcy of rationalism when dealing with this subject. Rickert deals with part of the pragmatistic position under the head of what he calls 'Relativismus'. I cannot discuss his text here. Suffice it to say that his argumentation in that chapter is so feeble as to seem almost incredible in so generally able a writer.

---

## Truth

Michael Dummett

Frege held that truth and falsity are the references of sentences. Sentences cannot stand for propositions (what Frege calls ‘thoughts’), since the reference of a complex expression depends only on the reference of its parts; whereas if we substitute for a singular term occurring in a sentence another singular term with the same reference but a different sense, the sense of the whole sentence, i.e., the thought which it expresses, changes. The only thing which it appears *must* in these circumstances remain unchanged is the truth-value of the sentence. The expressions “is true” and “is false” look like predicates applying to propositions, and one might suppose that truth and falsity were properties of propositions; but it now appears that the relation between a proposition and its truthvalue is not like that between a table and its shape, but rather like that between the sense of a definite description and the actual object for which it stands.

To the objection that there are non-truth-functional occurrences of sentences as parts of complex sentences, e.g., clauses in indirect speech, Frege replies that in such contexts we must take ordinary singular terms as standing, not for their customary reference, but for their sense, and hence we may say that in such a context, and only then, a sentence stands for the proposition it usually expresses.

If someone asks, “But what kind of entities are these truth-values supposed to be?” we may reply that there is no more difficulty in seeing what the truth-value of a sentence may be than there is in seeing what the direction of a line may be; we have been told when two sentences have the same truth-value—when they are materially equivalent—just as we know when two lines have the same direction—when they are parallel.



Nor need we waste time on the objection raised by Max Black that on Frege's theory certain sentences become meaningful which we should not normally regard as such, e.g., "If oysters are inedible, then the False." If sentences stand for truth-values, but there are also expressions standing for truth-values which are not sentences, then the objection to allowing expressions of the latter kind to stand wherever sentences can stand and vice versa is grammatical, not logical. We often use the word "thing" to provide a noun where grammar demands one and we have only an adjective, e.g., in "That was a disgraceful thing to do"; and we could introduce a verb, say "trues," to fulfil the purely grammatical function of converting a noun standing for a truth-value into a sentence standing for the same truth-value. It may be said that Frege has proved that a sentence does not ordinarily stand for a proposition, and has given a plausible argument that *if* sentences have references, they stand for truth-values, but that he has done nothing to show that sentences do have references at all. This is incorrect; Frege's demonstration that the notions of a concept (property) and a relation can be explained as special cases of the notion of a function provides a plausible argument for saying that sentences have a reference.

What *is* questionable is Frege's use of the words "truth" and "falsity" as names of the references of sentences; for by using these words rather than invented words of his own he gives the impression that by taking sentences to have a reference, with material equivalence as the criterion of identity, he has given an account of the notions of truth and falsity which we are accustomed to employ. Let us compare truth and falsity with the winning and losing of a board game. For a particular game we may imagine first formulating the rules by specifying the initial position and the permissible moves; the game comes to an end when there is no permissible move. We may then distinguish between two (or three) kinds of final positions, which we call "Win" (meaning that the player to make the first move wins), "Lose" (similarly), and, possibly, "Draw." Unless we tacitly appeal to the usual meanings of the words "win," "lose" and "draw," this description leaves out one vital point—that it is the object of a player to win. It is part of the concept of winning a game that a player plays to win, and this part of the concept is not conveyed by a classification of the end positions into winning ones and losing ones. We

can imagine a variant of chess in which it is the object of each player to be checkmated, and this would be an entirely different game; but the formal description we imagined would coincide with the formal description of chess. The whole theory of chess could be formulated with reference only to the formal description; but which theorems of this theory interested us would depend upon whether we wished to play chess or the variant game. Likewise, it is part of the concept of truth that we aim at making true statements; and Frege's theory of truth and falsity as the references of sentences leaves this feature of the concept of truth quite out of account. Frege indeed tried to bring it in afterwards, in his theory of assertion—but too late; for the sense of the sentence is not given in advance of our going in for the activity of asserting, since otherwise there could be people who expressed the same thoughts but went in instead for denying them.

A similar criticism applies to many accounts of truth and falsity or of the meanings of certain sentences in terms of truth and falsity. We cannot in general suppose that we give a proper account of a concept by describing those circumstances in which we do, and those in which we do not, make use of the relevant word, by describing the *usage* of that word; we must also give an account of the *point* of the concept, explain what we use the word *for*. Classifications do not exist in the void, but are connected always with some interest which we have, so that to assign something to one class or another will have consequences connected with this interest. A clear example is the problem of justifying a form of argument, deductive or inductive. Classification of arguments into (deductively or inductively) valid and invalid ones is not a game played merely for its own sake, although it *could* be taught without reference to any purpose or interest, say as a school exercise. Hence there is really a problem of showing that the criteria we employ for recognizing valid arguments do in fact serve the purpose we intend them to serve: the problem is not to be dismissed—as it has long been fashionable to do—by saying that we use the criteria we use.

We cannot assume that a classification effected by means of a predicate in use in a language will always have just *one* point. It may be that the classification of statements into true ones, false ones, and, perhaps, those that are neither true nor false, has one principal point, but that other

subsidiary ends are served by it which make the use of the words “true” and “false” more complex than it would otherwise be. At one time it was usual to say that we do not call ethical statements ‘true’ or ‘false,’ and from this many consequences for ethics were held to flow. But the question is not whether these words are in practice applied to ethical statements, but whether, if they were so applied, the point of doing so would be the same as the point of applying them to statements of other kinds, and, if not, in what ways it would be different. Again, to be told that we say of a statement containing a singular term which lacks reference that it is neither true nor false is so far only to be informed of a point of usage; no philosophical consequences can yet be drawn. Rather, we need to ask whether describing such a statement as neither true nor false accords better with the general point of classifying statements as true or false than to describe it as false. Suppose that we learn that in a particular language such statements are described as ‘false’: how are we to tell whether this shows that they use such statements differently from ourselves or merely that “false” is not an exact translation of their word? To say that we use singular statements in such a way that they are neither true nor false when the subject has no reference is meant to characterize our use of singular statements; hence it ought to be possible to describe when in a language not containing words for “true” and “false” singular statements would be sued in the same way as we use them, and when they would be used so as to be false when the subject had no reference. Until we have an account of the general point of the classification into true and false we do not know what interest attaches to saying of certain statements that they are neither true nor false; and until we have an account of how the truth-conditions of a statement determine its meaning the description of the meaning by stating the truth-conditions is valueless.

A popular account of the meaning of the word “true,” also deriving from Frege, is that  $\ulcorner$ It is true that  $P\urcorner$  has the same sense as the sentence  $P$ . If we then ask why it is any use to have the word “true” in the language, the answer is that we often refer to propositions indirectly, i.e., without expressing them, as when we say “Goldbach’s conjecture” or “what the witness said.” We also generalize about propositions without referring to any particular one, e.g., in “Everything he says is true.” This explanation cannot rank as a definition in the strict sense, since it permits elimination

of “is true” only when it occurs attached to a “that”-clause, and not when attached to any other expression standing for a proposition or to a variable; but, since every proposition can be expressed by a sentence, this does not refute its claim to be considered as determining uniquely the sense of “is true.” It might be compared with the recursive definition of “+,” which enables us to eliminate the sign “+” only when it occurs in front of a numeral, and not when it occurs in front of any other expression for a number or in front of a variable; yet there is a clear mathematical sense in which it specifies uniquely what operation “+” is to signify. Similarly, our explanation of “is true” determines uniquely the sense, or at least the application, of this predicate: for any given proposition there is a sentence expressing that proposition, and that sentence states the conditions under which the proposition is true.

If, as Frege thought, there exist sentences which express propositions but are neither true nor false, then this explanation appears incorrect. Suppose that  $P$  contains a singular term which has a sense but no reference: then, according to Frege,  $P$  expresses a proposition which has no truth-value. This proposition is therefore not true, and hence the statement  $\ulcorner$ It is true that  $P\urcorner$  will be *false*.  $P$  will therefore not have the same sense as  $\ulcorner$ It is true that  $P\urcorner$ , since the latter is false while the former is not. It is not possible to plead that  $\ulcorner$ It is true that  $P\urcorner$  is itself neither true nor false when the singular term occurring in  $P$  lacks a reference, since the *oratio obliqua* clause  $\ulcorner$ that  $P\urcorner$  stands for the proposition expressed by  $P$ , and it is admitted that  $P$  does have a sense and express a proposition; the singular term occurring in  $P$  has in  $\ulcorner$ It is true that  $P\urcorner$  its indirect reference, namely its sense, and we assumed that it did have a sense. In general, it will always be inconsistent to maintain the truth of every instance of “It is true that  $p$  if and only if  $p$ ” while allowing that there is a type of sentence which under certain conditions is neither true nor false. It would be possible to evade this objection by claiming that the “that”-clause in a sentence beginning “It is true that” is not an instance of *oratio obliqua*; that the word “that” here serves the purely grammatical function of transforming a sentence into a nounclause without altering either its sense or its reference. We should then have to take phrases like “Goldbach’s conjecture” and “what the witness said” as standing not for propositions but for truth-values. The expression “is true” would then be exactly like

the verb “trues” which we imagined earlier; it would simply convert a noun-phrase standing for a truth-value into a sentence without altering its sense or its reference. It might be objected that this variant of Frege’s account tallies badly with his saying that it is the *thought* (proposition) which is what is true or false; but we can express this point of Frege’s by saying that it is the *thought*, rather than the *sentence*, which primarily stands for a truthvalue. A stronger objection to the variant account is that it leans heavily on the theory of truth-values as references of sentences, while the original version depends only on the more plausible view that clauses in indirect speech stand for propositions. In any case, if there are meaningful sentences which say nothing which is true or false, then there must be *a* use of the word “true” which applies to propositions; for if we say  $\ulcorner$ It is neither true nor false that  $P\urcorner$ , the clause  $\ulcorner$ that  $P\urcorner$  must here be in *oratio obliqua*, otherwise the whole sentence would lack a truth-value.

Even if we do not wish to say of certain statements that they are neither true nor false, this account cannot give the *whole* meaning of the word “true.” If we are to give an explanation of the word “false” parallel to our explanation of “true” we shall have to say that  $\ulcorner$ It is false that  $P\urcorner$  has the same sense as the negation of  $P$ . In logical symbolism there exists a sign which, put in front of a sentence, forms the negation of that sentence; but in natural languages we do not have such a sign. We have to think to realize that the negation of “No-one is here” is not “No-one is not here” but “Someone is here”; there is no one rule for forming the negation of a given sentence. Now according to what principle do we recognize one sentence as the negation of another? It is natural to answer: The negation of a sentence  $P$  is that sentence which is true if and only if  $P$  is false and false if and only if  $P$  is true. But this explanation is ruled out if we want to use the notion of the negation of a sentence in order to explain the sense of the word “false.” It would not solve the difficulty if we did have a general sign of negation analogous to the logical symbol, for the question would then be: How in general do we determine the sense of the negation, given the sense of the original sentence?

We encounter the same difficulty over the connective “or.” We can give an account of the meaning of “and” by saying that we are in a position to assert  $\ulcorner P$  and  $Q\urcorner$  when and only when we are in a position to

assert P and in a position to assert Q. (This is not circular: one could train a dog to bark only when a bell rang *and* a light shone without presupposing that it possessed the concept of conjunction.) But, if we accept a two-valued logic, we cannot give a similar explanation of the meaning of “or.” We often assert  $\lceil P \text{ or } Q \rceil$  when we are not either in a position to assert P or in a position to assert Q. I use the word “we” here, meaning mankind, advisedly. If the history master gives the schoolboy a hint, saying, “It was either James I or Charles I who was beheaded,” then the schoolboy is in a position to assert, “Either James I or Charles I was beheaded” without (perhaps) being in a position to assert either limb of the disjunction; but it is not this sort of case which causes the difficulty. The *ultimate* source of the schoolboy’s knowledge derives from something which justifies the assertion that Charles I was beheaded; and this is all that would be required for the proposed explanation of the word “or” to be adequate. Likewise, the explanation is not impugned by cases like that in which I remember that I was talking either to Jean or to Alice, but cannot remember which. My knowledge that I was talking either to Jean or to Alice derives ultimately from the knowledge that I had at the time that I was talking to (say) Jean; the fact that the incomplete knowledge is all that survives is beside the point. Rather, the difficulty arises because we often make statements of the form  $\lceil P \text{ or } Q \rceil$  when the ultimate evidence for making them, in the sense indicated, is neither evidence for the truth of P nor evidence for the truth of Q. The most striking instance of this is the fact that we are prepared to assert *any* statement of the form  $\lceil P \text{ or not } P \rceil$ , even though we may have no evidence either for the truth of P or for the truth of  $\lceil \text{Not } P \rceil$ .

In order to justify asserting  $\lceil P \text{ or not } P \rceil$ , we appeal to the truth-table explanation of the meaning of “or.” But if the whole explanation of the meanings of “true” and “false” is given by “It is true that p if and only if p” and “It is false that p if and only if not p,” this appeal fails. The truth-table tells us, e.g., that from P we may infer  $\lceil P \text{ or } Q \rceil$  (in particular,  $\lceil P \text{ or not } P \rceil$ ); but *that* much we already knew from the explanation of “or” which we have rejected as insufficient. The truth-table does not show us that we are entitled to assert  $\lceil P \text{ or not } P \rceil$  in every possible case, since this is to assume that every statement is either true or false; but, if our explanation of “true” and “false” is all the explanation that can be given, to

say that every statement is either true or false is just to say that we are always justified in saying  $\lceil P \text{ or not } P \rceil$ .

We naturally think of truth-tables as giving the explanation of the sense which we attach to the sign of negation and to the connectives, an explanation which will show that we are justified in regarding certain forms of statement as logically true. It now appears that if we accept the redundancy theory of "true" and "false"—the theory that our explanation gives the whole meaning of these words—the truth-table explanation is quite unsatisfactory. More generally, we must abandon the idea which we naturally have that the notions of truth and falsity play an essential role in any account either of the meaning of statements in general or of the meaning of a particular statement. The conception pervades the thought of Frege that the general form of explanation of the sense of a statement consists in laying down the conditions under which it is true and those under which it is false (or better: saying that it is false under all other conditions); this same conception is expressed in the *Tractatus* in the words, "In order to be able to say that '*p*' is true (or false), I must have determined under what conditions I call '*p*' true, and this is how I determine the sense of the sentence" (4.063). But in order that someone should gain from the explanation that *P* is true in such-and-such circumstances an understanding of the sense of *P*, he must already know what it means to say of *P* that it is true. If when he inquires into this he is told that the only explanation is that to say that *P* is true is the same as to assert *P*, it will follow that in order to understand what is meant by saying that *P* is true, he must already know the sense of asserting *P*, which was precisely what was supposed to be being explained to him.

We thus have either to supplement the redundancy theory or to give up many of our preconceptions about truth and falsity. It has become a commonplace to say that there cannot be a criterion of truth. The argument is that we determine the sense of a sentence by laying down the conditions under which it is true, so that we could not first know the sense of a sentence and then apply some criterion to decide in what circumstances it was true. In the same sense there could not be a criterion for what constitutes the winning of a game, since learning what constitutes winning it is an essential part of learning what the game is. This does not mean that there may not be in any sense a theory of truth. For a

particular bounded language, if it is free of ambiguity and inconsistency, it must be possible to characterize the true sentences of the language; somewhat as, for a given game, we can say which moves are winning moves. (A language is bounded if we may not introduce into it new words or new senses for old words.) Such a characterization would be recursive, defining truth first for the simplest possible sentences, and then for sentences built out of others by the logical operations employed in the language; this is what is done for formalized languages by a truth-definition. The redundancy theory gives the general form of such a truth-definition, though in particular cases more informative definitions might be given.

Now we have seen that to say for each particular game what winning it consists in is not to give a satisfactory account of the concept of winning a game. What makes us use the same term "winning" for each of these various activities is that the point of every game is that each player tries to do what for that game constitutes winning; i.e., what constitutes winning always plays the same part in determining what playing the game consists in. Similarly, what the truth of a statement consists in always plays the same role in determining the sense of that statement, and a theory of truth must be possible in the sense of an account of what that role is. I shall not now attempt such an account; I claim, however, that such an account would justify the following. A statement, so long as it is not ambiguous or vague, divides all possible states of affairs into just *two* classes. For a given state of affairs, either the statement is used in such a way that a man who asserted it but envisaged that state of affairs as a possibility would be held to have spoken misleadingly, or the assertion of the statement would not be taken as expressing the speaker's exclusion of that possibility. If a state of affairs of the first kind obtains, the statement is false; if all actual states of affairs are of the second kind, it is true. It is thus *prima facie* senseless to say of any statement that in such-and-such a state of affairs it would be neither true nor false.

The sense of a statement is determined by knowing in what circumstances it is true and in what false. Likewise the sense of a command is determined by knowing what constitutes obedience to it and what disobedience; and the sense of a bet by knowing when the bet is won and when it is lost. Now there may be a gap between the winning of a bet



and the losing of it, as with a conditional bet; can there be a similar gap between obedience and disobedience to a command, or between the truth and falsity of a statement? There is a distinction between a conditional bet and a bet on the truth of a material conditional; if the antecedent is unfulfilled, in the first case the bet is off—it is just as if no bet had been made—but in the second case the bet is won. A conditional command where the antecedent is in the power of the person given the order (e.g., a mother says to a child, “If you go out, wear your coat”) is always like a bet on the material conditional; it is equivalent to the command to ensure the truth of the material conditional, viz., “Do not go out without your coat.” We cannot say that if the child does not go out, it is just as if no command had been given, since it may be that, unable to find his coat, he stayed in in order to comply with the command.

Can a distinction parallel to that for bets be drawn for conditional commands where the antecedent is not in the person’s power? I contend that the distinction which looks as if it could be drawn is in fact void of significance. There are two distinct kinds of consequence of making a bet, winning it and losing; to determine what is to involve one of these is not yet to determine completely what is to involve the other. But there is only one kind of consequence of giving a command, namely that, provided one had the right to give it in the first place, one acquires a right to punish or at least reprobate disobedience. It might be thought that punishment and reward were distinct consequences of a command in the same sense that paying money and receiving it are distinct consequences of a bet; but this does not tally with the role of commands in our society. The right to a reward is not taken to be an automatic consequence of obedience to a command, as the *right* to reproach is an automatic consequence of disobedience; if a reward is given, this is an act of grace, just as it is an act of grace if the punishment or reproach is withheld. Moreover, any action deliberately taken in order to comply with the command (to avoid disobedience to it) has the same claim to be rewarded as any other; hence to determine what constitutes disobedience to the command is thereby to determine what sort of behavior might be rewarded, without the need for any further decision. If the child stays in because he cannot find his coat, this behavior is as meritorious as if he goes out remembering to wear it; and if he forgets all about the order, but wears

his coat for some other reason, this behavior no more deserves commendation than if he chooses, for selfish reasons, to remain indoors. Where the antecedent is not in the person's power, it is indeed possible to regard the conditional command as analogous to the conditional bet; but since obedience to a command has no consequence of its own other than that of avoiding the punishment due for disobedience, there is not for such commands any significant distinction parallel to that between conditional bets and bets about a material conditional. If we regarded obedience to a command as giving a right to a reward, we could then introduce such a distinction for commands whose antecedent was in the person's power. Thus the mother might use the form, "If you go out, wear your coat," as involving that if the child went out with his coat he would be rewarded, if he went out without it he would be punished, and if he stayed indoors—even in order to comply with the command—he would be neither punished nor rewarded; while the form, "Do not go out without your coat," would involve his being rewarded if he stayed indoors.

Statements are like commands (as we use them) and not like bets; the making of a statement has, as it were, only one kind of consequence. To see this, let us imagine a language which contains conditional statements but has no counterfactual form (counterfactuals would introduce irrelevant complications). Two alternative accounts are suggested of the way in which conditionals are used in this language: one, that they are used to make statements conditionally; the other, that they represent the material conditional. On the first interpretation, a conditional statement is like a conditional bet: if the antecedent is fulfilled, then the statement is treated as if it had been an unconditional assertion of the consequent, and is said to be true or false accordingly; if the antecedent is not fulfilled, then it is just as if no statement, true or false, had been made at all. On the second interpretation, if the antecedent is not fulfilled, then the statement is said to be true. How are we to settle which of these two accounts is the correct one? If statements are really like bets and not like commands; if there are two distinct kinds of consequence which may follow the making of a statement, those that go with calling the statement 'true' and those that go with calling it 'false,' so that there may be a gap between these two kinds of consequence; then we ought to be able to find something which decides between the two accounts as definite as the financial transaction

which distinguishes a bet on the truth of the material conditional from a conditional bet. It is no use asking whether these people *say* that the man who has made a conditional statement whose antecedent turns out false said something true or that he said nothing true or false: they may have no words corresponding to “true” and “false”; and if they do, how could we be sure that the correspondence was exact? If their using the words “true” and “false” is to have the slightest significance, there must be some difference in their behavior which goes with their saying “true” or “neither true nor false” in this case.

It is evident on reflection that there is nothing in what they do which could distinguish between the two alternative accounts; the distinction between them is as empty as the analogous distinction for conditional commands whose antecedent is not in the person’s power. In order to fix the sense of an utterance, we do not need to make two separate decisions—when to say that a true statement has been made and when to say that a false statement has been made; rather, any situation in which nothing obtains which is taken as a case of its being false may be regarded as a case of its being true, just as someone who behaves so as not to disobey a command may be regarded as having obeyed it. The point becomes clearer when we look at it in the following way. If it makes sense in general to suppose that a certain form of statement is so used that in certain circumstances it is true, in others false, and in yet others nothing has been said true or false, then we can imagine that a form of conditional was used in this way (von Wright actually holds that *we* use conditionals in this way). If *P* turns out true, then “If *P*, then *Q*” is said to be true or false according as *Q* is true or false, while if *P* turns out false we say that nothing was said true or false. Let us contract this with what Frege and Strawson say about the use in our language of statements containing a singular term. If there is an object for which the singular term stands, then the statement is true or false according as the predicate does or does not apply to that object, but if there is no such object, then we have not said anything true or false. Now do these accounts tell us the sense of sentences of these two kinds?—that is, do they tell us how these statements are used, what is *done* by making statements of these forms? Not at all, for an essential feature of their use has not yet been laid down. Someone uttering a conditional statement of the kind described may very well have no opinion as to whether the antecedent was going to turn out

true or false; that is, he is not taken as having misused the statement or misled his hearers if he envisages it as a possibility that that case will arise in which he is said not to have made a statement true or false. All that he conveys by uttering the conditional statement is that he excludes the possibility that the case will arise in which he is said to have said something false, namely that antecedent is true and consequent false. With the case of a singular statement it is quite different. Here someone is definitely either misusing the form of statement or misleading his hearers if he envisages it as a possibility that that case will arise in which what he said will be said to be neither true nor false, namely that the singular term has no reference. He conveys more by making the statement than just that he excludes the possibility of its being false; he commits himself to its being true.

Are we then to say that laying down the truth-conditions for a sentence is not sufficient to determine its sense, that something further will have to be stipulated as well? Rather than say this we should abandon the notions of truth and falsity altogether. In order to characterize the sense of expressions of our two forms, only a twofold classification of possible relevant circumstances is necessary. We need to distinguish those states of affairs such that if the speaker envisaged them as possibilities he would be held to be either misusing the statement or misleading his hearers, and those of which this is not the case: and *one* way of using the words “true” and “false” would be to call states of affairs of the former kind those in which the statement was false and the others those in which the statement was true. For our conditional statements, the distinction would be between those states of affairs in which the statement was said to be false and those in which we said that it would either be true or else neither true nor false. For singular statements, the distinction would be between those states of affairs in which we said that the statement would either be false or else neither true nor false, and those in which it was true. To grasp the sense or use of these forms of statement, the twofold classification is quite sufficient; the threefold classification with which we started is entirely beside the point. Thus, on *one* way of using the words “true” and “false,” we should, instead of distinguishing between the conditional statement’s being true and its being neither true nor false, have distinguished between two different ways in which it could be true; and instead of distinguishing between the singular statement’s being false

and its being neither true nor false, we should have distinguished between two different ways in which it could be false.

This gives us a hint at a way of explaining the role played by truth and falsity in determining the sense of a statement. We have not yet seen what point there may be in distinguishing between different ways in which a statement may be true or between different ways in which it may be false, or, as we might say, between degrees of truth and falsity. The point of such distinctions does not lie in anything to do with the sense of the statement itself, but has to do with the way in which it enters into complex statements. Let us imagine that in the language of which the conditional statements we considered form a part there exists a sign of negation, i.e., a word which, placed in front of a statement, forms another statement; I call it a sign of negation because in most cases it forms a statement which we should regard as being used as the contradictory of the original statement. Let us suppose, however, that when placed in front of a conditional statement 'If P, then Q', it forms a statement which is used in the same way as the statement 'If P, then not Q'. Then if we describe the use of the conditionals by reference to a twofold classification only, i.e., in the same way as we describe a material conditional, we shall be unable to give a truth-functional account of the behavior of their sign "not." That is, we should have the tables:

P	Q	'If P, then Q'	'Not: if P, then Q'
T	T	T	F
T	F	F	T
F	T	T	T
F	F	T	T

in which the truth-value of 'Not: if P, then Q' is not determined by the truth-value of 'If P, then Q'. If, on the other hand, we revert to our original threefold classification, marking the case in which we said that no statement true or false had been made by "X," then we have the tables:

P	Q	'If P, then Q'	'Not: if P, then Q'
T	T	T	F
T	F	F	T
F	T	X	X
F	F	X	X

which can be quite satisfactorily accounted for by giving the table for “not”:

R	$\neg$ Not R
T	F
X	X
F	T

(I have assumed that the statements P and Q take only the values T and F.) It now becomes quite natural to think of “T” as representing “true,” “F” “false” and “X” “neither true nor false.” Then we can say that their symbol “not” really is a sign of negation, since  $\neg$ Not P is true when and only when P is false and false when and only when P is true. We must not forget, however, that the justification for distinguishing between the cases in which a conditional was said to have the value T and the cases in which it was said to have the value X was simply the possibility, created by this distinction, of treating “not” truth-functionally. In the same way if we have in a language an expression which normally functions as a sign of negation, but the effect of prefacing a singular statement with this expression is to produce a statement whose utterance still commits the speaker to there being an object for which the singular term stands, it is very natural to distinguish between two kinds of falsity a singular statement may have: that when the singular term has a reference, but the predicate does not apply to it, and that when the singular term lacks a reference. Let us represent the case in which the singular term has no reference by the symbol “Y,” and let us suppose S to be a singular statement. Then we have the table:

S	$\neg$ Not S
T	F
Y	Y
F	T

Here again it is natural to think of “T” as representing “true,” “F” “false” and “Y” “neither true nor false.”

There is no necessity to use the words “true” and “false” as suggested above, so that we have to interpret X as a kind of truth and Y as a kind of falsity. Logicians who study many-valued logics have a term which can be employed here: they would say that T and X are ‘designated’ truth-values and F and Y ‘undesignated’ ones. (In a many-valued logic those

formulas are considered valid which have a designated value for every assignment of values to their sentence-letters.) The points to observe are just these: (i) The sense of a sentence is determined wholly by knowing the case in which it has a designated value and the cases in which it has an undesignated one. (ii) Finer distinctions between different designated values or different undesignated ones, however naturally they come to us, are justified only if they are needed in order to give a truth-functional account of the formation of complex statements by means of operators. (iii) In *most* philosophical discussions of truth and falsity, what we really have in mind is the distinction between a designated and an undesignated value, and hence choosing the names “truth” and “falsity” for particular designated and undesignated values respectively will only obscure the issue. (iv) Saying that in certain circumstances a statement is neither true nor false does not determine whether the statement is in that case to count as having an undesignated or a designated value, i.e., whether someone who asserts the statement is or is not taken as excluding the possibility that that case obtains.

Baffled by the attempt to describe in general the relation between language and reality, we have nowadays abandoned the correspondence theory of truth, and justify our doing so on the score that it was an attempt to state a *criterion* of truth in the sense in which this cannot be done. Nevertheless, the correspondence theory expresses one important feature of the concept of truth which is not expressed by the law “It is true that p if and only if p” and which we have so far left quite out of account: that a statement is true only if there is something in the world *in virtue of which* it is true. Although we no longer accept the correspondence theory, we remain realists *au fond*; we retain in our thinking a fundamentally realist conception of truth. Realism consists in the belief that for any statement there must be something in virtue of which either it or its negation is true: it is only on the basis of this belief that we can justify the idea that truth and falsity play an essential role in the notion of the meaning of a statement, that the general form of an explanation of meaning is a statement of the truth-conditions.

To see the importance of this feature of the concept of truth, let us envisage a dispute over the logical validity of the statement “Either Jones was brave or he was not.” A imagines Jones to be a man, now dead, who

never encountered danger in his life. B retorts that it could still be true that Jones was brave, namely, if it is true that if Jones *had* encountered danger, he would have acted bravely. A agrees with this, but still maintains that it does not need to be the case that either “Jones was brave” = “If Jones had encountered danger, he would have acted bravely” nor “Jones was not brave” = “If Jones had encountered danger, he would not have acted bravely” is true. For, he argues, it might be the case that however many facts we knew of the kind which we should normally regard as grounds for asserting such counterfactual conditionals, we should still know nothing which would be a ground for asserting either. It is clear that B cannot agree that this is a possibility and yet continue to insist that all the same either “Jones was brave” or “Jones was not brave” is true; for he would then be committed to holding that a statement may be true even though there is nothing whatever such that, if we knew of it, we should count it as evidence or as a ground for the truth of the statement, and this is absurd. (It may be objected that there are assertions for which it would be out of place to ask one who made them for his evidence or grounds; but for *such* assertions the speaker must always either be in a position to make or in a position to deny them.) If B still wishes to maintain the necessity of “Either Jones was brave or he was not,” he will have to hold either that there must be some fact of the sort to which we usually appeal in discussing counterfactuals which, if we knew it, would decide us in favor either of the one counterfactual or of the other; or else that there is some fact of an extraordinary kind, perhaps known only to God. In the latter case he imagines a kind of spiritual mechanism—Jones’ character—which determines how he acts in each situation that arises; his acting in such-and-such a way reveals to us the state of this spiritual mechanism, which was however already in place before its observable effects were displayed in his behavior. B would then argue thus: If Jones *had* encountered danger, he would either have acted bravely or have acted like a coward. Suppose he had acted bravely. This would then have shown us that he was brave; but he would *already* have been brave before his courage was revealed by his behavior. That is, either his character included the quality of courage or it did not, and his character determines his behavior. We know his character only indirectly, through its effects on his behavior; but each character-trait



must be *there* within him independently of whether it reveals itself to us or not.

Anyone of a sufficient degree of sophistication will reject B's belief in a spiritual mechanism; either he will be a materialist and substitute for it an equally blind belief in a physiological mechanism, or he will accept A's conclusion that "Either Jones was brave or he was not" is not logically necessary. His ground for rejecting B's argument is that if such a statement as "Jones was brave" is true, it must be true in virtue of the sort of fact we have been taught to regard as justifying us in asserting it. It cannot be true in virtue of a fact of some quite different sort of which we can have no direct knowledge, for otherwise the statement "Jones was brave" would not have the meaning that *we* have given it. In accepting A's position he makes a small retreat from realism; he abandons a realist view of character.

In order, then, to decide whether a realist account of truth can be given for statements of some particular kind, we have to ask whether for such a statement P it must be the case that if we knew sufficiently many facts of the kind we normally treat as justifying us in asserting P, we should be in a position either to assert P or to assert  $\neg$  P: if so, then it can truly be said that there must either be something in virtue of which P is true or something in virtue of which it is false. It is easy to overlook the force of the phrase "sufficiently many." Consider the statement "A city will never be built on this spot." Even if we have an oracle which can answer every question of the kind, "Will there be a city here in 1990?" "In 2100?" etc., we might never be in a position either to declare the statement true or to declare it false. Someone may say: That is only because you are assuming the knowledge of only finitely many answers of the oracle; but if you knew the oracle's answers to *all* these questions, you would be able to decide the truth-value of the statement. But what would it mean to know infinitely many facts? It could mean that the oracle gave a direct answer "No" to the question, "Will a city ever be built here?": but to assume this is just like B's assumption of the existence of a hidden spiritual mechanism. It might mean that we had an argument to show the falsity of  $\neg$  "A city will be built here in the year N" irrespective of the value of N, e.g., if 'here' is the North Pole: but no one would suggest that it must be the case that either the oracle will give an affirmative answer to

some question of the form “Will there be a city here in the year . . . ?” or we can find a general argument for a negative answer. Finally, it could mean that we were *able* to answer every question of the form, “Will there be a city here in the year . . . ?”: but having infinite knowledge in *this* sense will place us in no better position than when we had the oracle.

We thus arrive at the following position. We are entitled to say that a statement P must be either true or false, that there must be something in virtue of which either it is true or it is false, only when P is a statement of such a kind that we could in a finite time bring ourselves into a position in which we were justified either in asserting or in denying P; that is, when P is an effectively decidable statement. This limitation is not trivial: there is an immense range of statements which, like “Jones was brave,” are concealed conditionals, or which, like “A city will never be built here,” contain—explicitly or implicitly—an unlimited generality, and which therefore fail the test.

What I have done here is to transfer to ordinary statements what the intuitionists say about mathematical statements. The sense of e.g., the existential quantifier is determined by considering what sort of fact makes an existential statement true, and this means: the sort of fact which we have been taught to regard as justifying us in asserting an existential statement. What would make the statement that there exists an odd perfect number true would be some particular number’s being both odd and perfect; hence the assertion of the existential statement must be taken as a claim to be able to assert some one of the singular statements. We are thus justified in asserting that there is a number with a certain property only if we have a method for finding a particular number with that property. Likewise, the sense of a universal statement is given by the sort of consideration we regard as justifying us in asserting it: namely we can assert that every number has a certain property if we have a general method for showing, for any arbitrary number, that it has that property. Now what if someone insists that either the statement “There is an odd perfect number” is true, or else every perfect number is even? He is justified if he knows of a procedure which will lead him in a finite time either to the determination of a particular odd perfect number or to a general proof that a number assumed to be perfect is even. But if he knows of no such procedure, then he is trying to attach to the statement “Every per-

fect number is even" a meaning which lies *beyond* that provided by the training we are given in the use of universal statements; he wants to say, as B said of "Jones was brave," that its truth may lie in a region directly accessible only to God, which human beings can never survey.

We learn the sense of the logical operators by being trained to *use* statements containing them, i.e., to assert such statements under certain conditions. Thus we learn to assert  $\lceil P \text{ and } Q \rceil$  when we can assert  $P$  and can assert  $Q$ ; to assert  $\lceil P \text{ or } Q \rceil$  when we can assert  $P$  or can assert  $Q$ ; to assert  $\lceil \text{For some } n, F(n) \rceil$  when we can assert  $\lceil F(0) \rceil$  or can assert  $\lceil F(1) \rceil$  or .... We learn to assert  $\lceil \text{For every } n, F(n) \rceil$  when we can assert  $\lceil F(0) \rceil$  and  $\lceil F(1) \rceil$  and ...; and to say that we can assert all of these means that we have a general method for establishing  $\lceil F(x) \rceil$  irrespective of the value of  $x$ . Here we have abandoned altogether the attempt to explain the meaning of a statement by laying down its truth-conditions. *We no longer explain the sense of a statement by stipulating its truth-value in terms of the truth-values of its constituents, but by stipulating when it may be asserted in terms of the conditions under which its constituents may be asserted.* The justification for this change is that this is how we in fact learn to use these statements: furthermore, the notions of truth and falsity cannot be satisfactorily explained so as to form a basis for an account of meaning once we leave the realm of effectively decidable statements. One result of this shift in our account of meaning is that, unless we are dealing only with effectively decidable statements, certain formulas which appeared in the two-valued logic to be logical laws no longer rank as such, in particular the law of excluded middle: this is rejected, not on the ground that there is a middle truth-value, but because meaning, and hence validity, is no longer to be explained in terms of truth-values.

Intuitionists speak of mathematics in a highly antirealist (antiplatonic) way: for them it is *we* who construct mathematics; it is not already *there* waiting for us to discover. An extreme form of such constructivism is found in Wittgenstein's *Remarks on the Foundations of Mathematics*. This makes it appear as though the intuitionist rejection of an account of the meaning of mathematical statements in terms of truth and falsity could not be generalized for other regions of discourse, since even if there is no independent mathematical reality answering to our mathematical

statements, there is an independent reality answering to statements of other kinds. On the other hand the exposition of intuitionism I have just given was not based on a rejection of the Fregean notion of a mathematical reality waiting to be discovered, but only on considerations about meaning. Now certainly someone who accepts the intuitionist standpoint in mathematics will not be inclined to adopt the platonist picture. Must he then go to the other extreme, and have the picture of our creating mathematics as we go along? To adopt this picture involves thinking with Wittgenstein that we are *free* in mathematics at every point; no step we take has been forced on us by a necessity external to us, but has been freely chosen. This picture is not the only alternative. If we think that mathematical results are in some sense imposed on us from without, we could have instead the picture of a mathematical reality not already in existence but as it were coming into being as we probe. Our investigations bring into existence what was not there before, but what they bring into existence is not of our own making.

Whether this picture is right or wrong for mathematics, it is available for other regions of reality as an alternative to the realist conception of the world. This shows how it is possible to hold that the intuitionist substitution of an account of the *use* of a statement for an account of its truth-conditions as the general form of explanation of meaning should be applied to all realms of discourse without thinking that we create the world; we can abandon realism without falling into subjective idealism. This substitution does not, of course, involve dropping the words “true” and “false,” since for most ordinary contexts the account of these words embodied in the laws “It is true that  $p$  if and only if  $p$ ” and “It is false that  $p$  if and only if not  $p$ ” is quite sufficient: but it means facing the consequences of admitting that this is the *whole* explanation of the sense of these words, and this involves dethroning truth and falsity from their central place in philosophy and in particular in the theory of meaning. Of course the doctrine that meaning is to be explained in terms of use is the cardinal doctrine of the later Wittgenstein; but I do not think the point of this doctrine has so far been generally understood.



---

## Two Philosophical Perspectives

Hilary Putnam

The problems we have been discussing naturally give rise to two philosophical points of view (or two philosophical temperaments, as I called them in the Introduction). It is with these points of view, and with their consequences for just about every issue in philosophy that I shall be concerned: the question of 'Brains in a Vat' would not be of interest, except as a sort of logical paradox, if it were not for the sharp way in which it brings out the difference between these philosophical perspectives.

One of these perspectives is the perspective of metaphysical realism. On this perspective, the world consists of some fixed totality of mind-independent objects. There is exactly one true and complete description of 'the way the world is'. Truth involves some sort of correspondence relation between words or thought-signs and external things and sets of things. I shall call this perspective the *externalist* perspective, because its favorite point of view is a God's Eye point of view.

The perspective I shall defend has no unambiguous name. It is a late arrival in the history of philosophy, and even today it keeps being confused with other points of view of a quite different sort. I shall refer to it as the *internalist* perspective, because it is characteristic of this view to hold that *what objects does the world consist of?* is a question that it only makes sense to ask *within* a theory or description. Many 'internalist' philosophers, though not all, hold further that there is more than one 'true' theory or description of the world. 'Truth', in an internalist view, is some sort of (idealized) rational acceptability—some sort of ideal coherence of our beliefs with each other and with our experiences *as those experiences are themselves represented in our belief system*—and not correspondence with mind-independent or discourse-independent 'states

of affairs'. There is no God's Eye point of view that we can know or usefully imagine; there are only the various points of view of actual persons reflecting various interests and purposes that their descriptions and theories subserve. ('Coherence theory of truth'; 'Non-realism'; 'Verificationism'; 'Pluralism'; 'Pragmatism'; are all terms that have been applied to the internalist perspective; but every one of these terms has connotations that are unacceptable because of their other historic applications.)

Internalist philosophers dismiss the 'Brain in a Vat' hypothesis. For us, the 'Brain in a Vat World' is only a *story*, a mere linguistic construction, and not a possible world at all. The idea that this story might be true in some universe, some Parallel Reality, assumes a God's Eye point of view from the start, as is easily seen. For *from whose point of view is the story being told?* Evidently *not* from the point of view of any of the sentient creatures *in* the world. Nor from the point of view of any observer in another world who interacts with this world; for a 'world' by definition includes everything that interacts in any way with the things it contains. If *you*, for example, were the one observer who was *not* a Brain in a Vat, spying on the Brains in a Vat, then the world would not be one in which *all* sentient beings were Brains in a Vat. So the supposition that there could be a world in which *all* sentient beings are Brains in a Vat presupposes from the outset a God's Eye view of truth, or, more accurately, a No Eye view of truth—truth as independent of observers altogether.

For the externalist philosopher, on the other hand, the hypothesis that we are all Brains in a Vat cannot be dismissed so simply. For the truth of a theory does not consist in its fitting the world as the world presents itself to some observer or observers (truth is not 'relational' in this sense), but in its corresponding to the world as it is in itself. And the problem that I posed for the externalist philosopher is that the very relation of correspondence on which truth and reference depend (on his view) cannot logically be available to him if he *is* a Brain in a Vat. So, if we *are* Brains in a Vat, we cannot *think* that we are, except in the bracketed sense [we are Brains in a Vat]; and this bracketed thought does not have reference conditions that would make it *true*. So it is not possible after all that we are Brains in a Vat.

Suppose we assume a 'magical theory of reference'. For example, we might assume that some occult rays—call them 'noetic rays'<sup>1</sup>—connect

words and thought-signs to their referents. Then there is no problem. The Brain in a Vat can think the *words*, 'I am a brain in a vat', and when he does the word 'vat' corresponds (with the aid of the noetic rays) to real external vats and the word 'in' corresponds (with the aid of the noetic rays) to the relation of real spatial containment. But such a view is obviously untenable. No present day philosopher would espouse such a view. It is because the modern realist wishes to have a correspondence theory of truth *without* believing in 'noetic rays' (or, believing in Self-Identifying Objects<sup>2</sup>—objects that intrinsically correspond to one word or thought-sign rather than another) that the Brain in a Vat case is a puzzler for him.

As we have seen, the problem is this: there are these objects out there. Here is the mind/brain, carrying on its thinking/computing. How do the thinker's symbols (or those of his mind/brain) get into a unique correspondence with objects and sets of objects out there?

The reply popular among externalists today is that while indeed no sign *necessarily* corresponds to one set of things rather than another, *contextual* connections between signs and external things (in particular, causal connections) will enable one to explicate the nature of reference. But this doesn't work. For example, the dominant cause of my beliefs about electrons is probably various *textbooks*. But the occurrences of the word 'electron' I produce, though having in this sense a strong connection to textbooks, do not *refer* to textbooks. The objects which are the dominant cause of my beliefs containing a certain sign may not be the referents of that sign.

The externalist will now reply that the word 'electron' is not connected to textbooks by a causal chain *of the appropriate type*. (But how can we have intentions which determine which causal chains are 'of the appropriate type' unless we are *already* able to *refer*?)

For an internalist like myself, the situation is quite different. In an internalist view also, signs do not intrinsically correspond to objects, independently of how those signs are employed and by whom. But a sign that is actually employed in a particular way by a particular community of users can correspond to particular objects *within the conceptual scheme of those users*. 'Objects' do not exist independently of conceptual schemes. *We* cut up the world into objects when we introduce one or



another scheme of description. Since the objects *and* the signs are alike *internal* to the scheme of description, it is possible to say what matches what.

Indeed, it is trivial to say what any word refers to *within* the language the word belongs to, by using the word itself. What does 'rabbit' refer to? Why, to rabbits, of course! What does 'extraterrestrial' refer to? To extraterrestrials (if there are any).

Of course the externalist agrees that the extension of 'rabbit' is the set of rabbits and the extension of 'extraterrestrial' is the set of extraterrestrials. But he does not regard such statements as telling us what reference *is*. For him finding out what reference *is*, i.e. what the *nature* of the 'correspondence' between words and things is, is a pressing problem. (*How* pressing, we saw in the previous chapter.) For me there is little to say about what reference is within a conceptual system other than these tautologies. The idea that causal connection is necessary is refuted by the fact that 'extraterrestrial' certainly refers to extraterrestrials whether we have ever causally interacted with any extraterrestrials or not!

The externalist philosopher would reply, however, that we can refer to extraterrestrials even though we have never interacted with any (as far as we know) because we have interacted with *terrestrials* and we have experienced instances of the relation 'not from the same planet as' and instances of the property 'intelligent being'. And we can *define* an extraterrestrial as an intelligent being that is not from the same planet as terrestrials. Also, 'not from the same planet as' can be analyzed in terms of 'not from the same place as' and 'planet' (which can be further analyzed). Thus the externalist gives up the requirement that we have some 'real' connection (e.g. causal connection) with *everything* we are able to refer to, and requires only that the *basic* terms refer to kinds of things (and relations) that we have some real connection to. Using the basic terms in complex combinations we can then, he says, build up descriptive expressions which refer to kinds of things we have no real connection to, and that may not even exist (e.g. extraterrestrials).

In fact, already with a simple word like 'horse' or 'rabbit' he might have observed that the extension includes many things we have *not* causally interacted with (e.g. *future* horses and rabbits, or horses and rabbits that never interacted with any human being). When we use the

word 'horse' we refer not only to the horses we have a real connection to, but also to all other things *of the same kind*.

At this point, however, we must observe that 'of the same kind' makes no sense apart from a categorial system which says what properties do and what properties do not count as similarities. In *some* ways, after all, anything is 'of the same kind' as anything else. This whole complicated story about how we refer to some things by virtue of the fact that they are connected with us by 'causal chains of the appropriate kind', and to yet other things by virtue of the fact that they are 'of the same kind' as things connected with us by causal chains of the appropriate kind, and to still other things 'by description', is not so much false as otiose. What makes horses with which I have not interacted 'of the same kind' as horses with which I *have* interacted is that fact that the former as well as the latter are *horses*. The metaphysical realist formulation of the problem once again makes it seem as if there are to begin with all these objects in themselves, and then I get some kind of a lasso over a few of these objects (the horses with which I have a 'real' connection, via a 'causal chain of the appropriate kind'), and then I have the problem of getting my word ('horse') to cover not only the ones I have 'lassoed' but also the ones I can't lasso, because they are too far away in space and time, or whatever. And the 'solution' to this pseudo-problem, as I consider it to be—the metaphysical realist 'solution'—is to say that the word *automatically* covers not just the objects I lassoed, but also the objects which are *of the same kind*—of the same kind *in themselves*. But then the world is, after all, being claimed to contain Self-Identifying Objects, for this is just what it means to say that the *world*, and not thinkers, sorts things into kinds.

In a sense, I would say, the world *does* consist of 'Self-Identifying Objects'—but not a sense available to an externalist. If, as I maintain, 'objects' themselves are as much made as discovered, as much products of our conceptual invention as of the 'objective' factor in experience, the factor independent of our will, then of course objects intrinsically belong under certain labels; because those labels are the tools we used to construct a version of the world with such objects in the first place. But *this* kind of 'Self-Identifying Object' is not mind-independent; and the externalist wants to think of the world as consisting of objects that are *at*

*one and the same time* mind-independent and Self-Identifying. This is what one cannot do.

### Internalism and Relativism

Internalism is not a facile relativism that says, 'Anything goes'. Denying that it makes sense to ask whether our concepts 'match' something totally uncontaminated by conceptualization is one thing; but to hold that every conceptual system is therefore just as good as every other would be something else. If anyone really believed that, and if they were foolish enough to pick a conceptual system that told them they could fly and to act upon it by jumping out of a window, they would, if they were lucky enough to survive, see the weakness of the latter view at once. Internalism does not deny that there are experiential *inputs* to knowledge; knowledge is not a story with no constraints except *internal* coherence; but it does deny that there are any inputs *which are not themselves to some extent shaped by our concepts*, by the vocabulary we use to report and describe them, or any inputs *which admit of only one description, independent of all conceptual choices*. Even our description of our own sensations, so dear as a starting point for knowledge to generations of epistemologists, is heavily affected (as are the sensations themselves, for that matter) by a host of conceptual choices. The very inputs upon which our knowledge is based are conceptually contaminated; but contaminated inputs are better than none. If contaminated inputs are all we have, still all we have has proved to be quite a bit.

What makes a statement, or a whole system of statements—a theory or conceptual scheme—rationally acceptable is, in large part, its coherence and fit; coherence of 'theoretical' or less experiential beliefs with one another and with more experiential beliefs, and also coherence of experiential beliefs with theoretical beliefs. Our conceptions of coherence and acceptability are, on the view I shall develop, deeply interwoven with our psychology. They depend upon our biology and our culture; they are by no means 'value free'. But they *are* our conceptions, and they are conceptions of something real. They define a kind of objectivity, *objectivity for us*, even if it is not the metaphysical objectivity of the God's Eye view.

Objectivity and rationality humanly speaking are what we have; they are better than nothing.

To reject the idea that there is a coherent 'external' perspective, a theory which is simply true 'in itself', apart from all possible observers, is not to *identify* truth with rational acceptability. Truth cannot simply *be* rational acceptability for one fundamental reason; truth is supposed to be a property of a statement that cannot be lost, whereas justification can be lost. The statement 'The earth is flat' was, very likely, rationally acceptable 3,000 years ago; but it is not rationally acceptable today. Yet it would be wrong to say that 'the earth is flat' was *true* 3,000 years ago; for that would mean that the earth has changed its shape. In fact, rational acceptability is both tensed and relative to a person. In addition, rational acceptability is a matter of degree; truth is sometimes spoken of as a matter of degree (e.g., we sometimes say, '*the earth is a sphere*' is *approximately true*); but the 'degree' here is the *accuracy* of the statement, and not its degree of acceptability or justification.

What this shows, in my opinion, is not that the externalist view is right after all, but that truth is an *idealization* of rational acceptability. We speak as if there were such things as epistemically ideal conditions, and we call a statement 'true' if it would be justified under such conditions. 'Epistemically ideal conditions', of course, are like 'frictionless planes': we cannot really attain epistemically ideal conditions, or even be absolutely certain that we have come sufficiently close to them. But frictionless planes cannot really be attained either, and yet talk of frictionless planes has 'cash value' because we can approximate them to a very high degree of approximation.

Perhaps it will seem that explaining truth in terms of justification under ideal conditions is explaining a clear notion in terms of a vague one. But 'true' is *not* so clear when we move away from such stock examples as 'Snow is white.' And in any case, I am not trying to give a formal *definition* of truth, but an informal elucidation of the notion.

The simile of frictionless planes aside, the two key ideas of the idealization theory of truth are (1) that truth is independent of justification here and now, but not independent of *all* justification. To claim a statement is true is to claim it could be justified. (2) truth is expected to be

stable or 'convergent'; if both a statement and its negation could be 'justified', even if conditions were as ideal as one could hope to make them, there is no sense in thinking of the statement as *having* a truth-value.

#### Notes

1. 'Noetic rays' was suggested to me by Zemach.
2. The term 'Self Identifying Object' is from *Substance and Sameness* by David Wiggins (Blackwell, 1980).

---

## Is Truth a Goal of Inquiry? Donald Davidson versus Crispin Wright

Richard Rorty

Pragmatists think that if something makes no difference to practice, it should make no difference to philosophy. This conviction makes them suspicious of the distinction between justification and truth, for that difference makes no difference to my decisions about what to do. If I have concrete, specific doubts about whether one of my beliefs is true, I can resolve those doubts only by asking whether it is adequately justified—by finding and assessing additional reasons pro and con. I cannot bypass justification and confine my attention to truth: assessment of truth and assessment of justification are, when the question is about what I should believe now, the same activity.<sup>1</sup> If, on the other hand, my doubts are as unspecific and abstract as Descartes's—are such that I can do nothing to resolve them—they should be dismissed, with Peirce, as “make-believe.” Philosophy should ignore them.

This line of thought suggests to pragmatists that, although there is obviously a lot to be said about justification of various sorts of beliefs, there may be little to say about truth.<sup>2</sup> The sort of thing philosophers typically have said—that truth is some sort of correspondence to, or accurate representation of, reality—seemed empty and pointless to many<sup>3</sup> nineteenth-century idealists, and also to Dewey. The early pragmatists agreed with their idealist opponents that doubts about correspondence to reality can be settled only by assessing the coherence of the dubious belief with other beliefs. To both, the difference between true beliefs considered as useful nonrepresentational mental states, and as accurate (and *therefore* useful) representations of reality, seemed a difference that could make no difference to practice. No one profits from insisting on the distinction, both concluded, except for those who enjoy entertaining make-believe doubts.

Since the pragmatists, unlike the idealists, took Darwin and biology seriously, they had an additional reason for distrusting the idea that true beliefs are accurate representations. For representation, as opposed to increasingly complex adaptive behavior, is hard to integrate into an evolutionary story. Within such a story, it is easy to think of beliefs, with Bain and Peirce, as habits of action, patterns of complex behavior. But it is hard to imagine that, at a certain point in the evolutionary process, somewhere between the squids and the apes, these patterns began to be determined by inner representations, having previously been determined by mere neurological configurations. Even if one chooses to treat sufficiently complex neurological configurations *as* representations, the question of their accuracy seems to collapse immediately into that of their utility. So, once again, we seem to have a difference that makes no practical difference.<sup>4</sup>

William James said, “‘The true’ . . . is only the expedient in the way of our thinking, just as ‘the right’ is only the expedient in the way of our behaving.”<sup>5</sup> Elsewhere he said, “The true is the name of whatever proves itself to be good in the way of belief, and good, too, for definite, assignable reasons.”<sup>6</sup> His point in analogizing truth to rightness and to goodness was that once you understand all about the justification of actions, including the justification of assertions, you understand all there *is* to understand about goodness, rightness, and truth.<sup>7</sup>

Philosophers who, like myself, find this Jamesian suggestion persuasive, swing back and forth between trying to reduce truth to justification and propounding some form of minimalism about truth. In reductionist moods we have offered such definitions of truth as “warranted assertibility,” “ideal assertibility,” and “assertibility at the end of inquiry.” But such definitions always fall victim, sooner or later, to what Putnam has called the “naturalistic fallacy” argument—the argument that a given belief might meet any such conditions but still not be true. Faced with this argument, we pragmatists have often fallen back on minimalism and have suggested that Tarski’s breezy disquotationalism may exhaust the topic of truth.<sup>8</sup>

In an article on Donald Davidson published in 1986, I suggested that we interpret Davidson both as a sort of pragmatist and as a sort of minimalist—as someone who, like James, thought that there was less

to say about truth than philosophers had usually believed.<sup>9</sup> More specifically, I interpreted Davidson as saying that the word “true” had no explanatory use, but merely a disquotational use, a commending use, and what I called a “cautionary” use. The latter is its use in such expressions as “fully justified, but perhaps not true.” The reason there is less to be said about truth than one might think, I suggested, is that terms used to commend or caution—terms such as “good!” “right!” “true!” “false!” “way to go!” and “watch it!”—do not need much philosophical definition or explication.

My underlying idea in that 1986 article was that the entire force of the cautionary use of “true” is to point out that justification is relative to an audience and that we can never exclude the possibility that some better audience might exist, or come to exist, to whom a belief that is justifiable to us would not be justifiable. But, as Putnam’s “naturalistic fallacy” argument shows, there can be no such thing as an “ideal audience” before which justification would be sufficient to ensure truth, any more than there can be a largest integer. For any audience, one can imagine a better-informed audience and also a more imaginative one—an audience that has thought up hitherto-undreamt-of alternatives to the proposed belief. The limits of justification would be the limits of language, but language (like imagination) has no limits.

In an article of 1990, Davidson partially repudiated my interpretation.<sup>10</sup> He said that he should be considered neither a deflationist nor a disquotationalist. He defined “deflationism” as the view that “Tarski’s work embraces all of truth’s essential features” and said that I was mistaken in attributing this view to him on the basis of his eschewal of attempts to define “true” for variable *L* as opposed to defining “true-in-*L*” for particular values of *L*.<sup>11</sup> He went on to say that

Tarski’s definitions [of the term “true-in-*L*” for various values of *L*] give us no idea of how to apply the concept [of truth] to a new case. . . . [T]hey depend on giving the extension or references of the basic predicates or names by enumerating cases; a definition given in this way can provide no clue for the next or general case.<sup>12</sup>

Davidson concluded that “[t]he concept of truth has essential connections with the concepts of belief and meaning, but these connections are untouched by Tarski’s work.”<sup>13</sup> He summed up by saying:



What Tarski has done for us is to show in detail how to describe the kind of pattern truth must make. What we need to do now is to say how to identify the presence of such a pattern or structure in the behavior of people.<sup>14</sup>

The way we identify this pattern, Davidson tells us, is to gather information “about what episodes and situations in the world cause an agent to prefer that one rather than another sentence be true.”<sup>15</sup> This information can be gleaned without knowing what the agent’s sentences mean. But once we have enough such evidence we can, Davidson says, “make the crucial step from the nonpropositional to the propositional,”<sup>16</sup> from the nonintensional to the intensional. For the use of intensional terms to describe human behavior marks the emergence of the pattern that truth makes—the pattern that links those episodes and situations in the world with the noises and marks made by the agent. They are linked into the behavior we call “using a language.” Detection of that pattern is what makes the adoption of what Dennett calls “the intentional stance” both possible and useful in our dealings with the agent. There is, Davidson says, “a fundamentally rational pattern that must, in general outline, be shared by all rational creatures.”<sup>17</sup> This pattern that rationality makes is the same pattern truth makes, and the same pattern meaning makes. You cannot have language without rationality, or either without truth.<sup>18</sup>

It is important to realize that what Davidson adds to Tarski, when he displays the connections between the concept of truth and those of meaning and belief, has nothing whatever to do with the question of whether, or how, we can tell when a belief is true. Although Davidson describes himself as, in his Dewey Lectures, filling in the missing “content” of the “concept” of truth, all this filling-in amounts to is instructions for constructing an empirical theory for explaining and predicting behavior—a theory of truth for one or more speakers. “A theory of truth,” as he says, “is an empirical theory about the truth conditions of every sentence in some corpus of sentences.”<sup>19</sup>

Philosophers who discuss truth have often hoped to underwrite our assumption that, the more justification we offer of a belief, the likelier it is that that belief is true. The most familiar attempt at such ratification begins by saying that, at least in some areas of culture, and at the very least when we are concerned with observable physical objects, our predictions succeed insofar as our beliefs fit reality. It then goes on to say

that each successive substitution of a better-justified for a worse-justified belief is an improvement in degree of fit. Such talk of “fit” interprets an increase in the coherence of nonobservational sentences with observation sentences as a sign of closer fit between the former sentences and the things observed.

Davidson, however, has no sympathy for this line of thought. His criticisms of the notion of “fitting reality,” in “On the Very Idea of a Conceptual Scheme,” parallel James’s and Dewey’s. In his Dewey Lectures he says:

I have argued that certain familiar attempts to characterize truth which go beyond giving empirical content to a structure of the sort Tarski taught us how to describe are empty, false, or confused. We should not say that truth is correspondence, coherence, warranted assertibility, ideally justified assertibility, what is accepted in the conversation of the right people, what science will end up maintaining, what explains the convergence on single theories in science, or the success of our ordinary beliefs. To the extent that realism or antirealism depend [*sic*] on one or another of these views of truth we should refuse to endorse either.<sup>20</sup>

Passages such as this suggest that Davidson would categorically repudiate the suggestion that philosophers need to explain why an increase in justification leads to an increased likelihood of truth, as opposed to acceptability to more and more audiences. For Davidson seems to think that philosophers have done *all* they need to do with the concept of truth once they have shown how to detect a certain pattern of behavior—the pattern exhibited in the truth theory for a language. It is hard to see how such detection could help to underwrite or improve our practices of justification, and Davidson gives no reason to think that it could or should.<sup>21</sup> This is, presumably, why he calls truth a “nonepistemic” concept.

I suspect that the only epistemological comfort that Davidson has to offer is his notorious thesis that most of our beliefs—most of *anybody’s* beliefs—must be true. This thesis is, however, both less bracing and less provocative than it may seem at first. For when we remember that Davidson will have no truck with the idea that truth consists in correspondence to, or accurate representation of,<sup>22</sup> reality, we realize that he is not saying that our minds are, thanks to God’s or Evolution’s contrivance, well suited to the task of getting reality right. He can perfectly well agree with Goodman, Putnam, and Kuhn that there is no such task, because there is no Way the World Is. He is, rather, saying that most

of anybody's beliefs must coincide with most of *our* beliefs (because to ascribe beliefs in the first place one must invoke the Principle of Charity) and that to reject that mass of shared beliefs (as perhaps not corresponding to reality) is to bring back a tangle of uncashable and useless metaphors—those used to state the scheme-content distinction. To say, as Davidson does, that “belief is in its nature veridical” is not to celebrate the happy congruence of subject and object but rather to say that the pattern truth makes is the pattern that *justification to us* makes.<sup>23</sup>

Without charity, we cannot detect the pattern truth makes. But charity entails seeing most of what the natives say as justified. If there is no justification of the sort that strikes us as reasonable, there will be no coherent set of inferential relationships to be detected between the various strings of marks and noises produced by speakers, and therefore no rationality—no pattern of the requisite sort. This seems to me the *sole* force of Davidson's claim that the guiding principles used in detecting this pattern “derive from normative considerations”<sup>24</sup> and of his reference to “the norms that govern our theories of intensional attribution.”<sup>25</sup> The need to justify our beliefs and desires to ourselves and to our fellow agents subjects us to norms, and obedience to these norms produces a behavioral pattern that we must detect in others before confidently attributing beliefs to them. But there seems no occasion to look for obedience to an *additional* norm—the commandment to seek the truth. For—to return to the pragmatist doubt with which I began—obedience to that commandment will produce no behavior not produced by the need to offer justification.

So far I have been sketching the sort of minimalism about truth that I would still wish to attribute to Davidson, even after accepting his repudiation of deflationism. But this minimalism is very different from certain other philosophical accounts of truth that have been called by that name. To highlight these differences, I turn now from Davidson to Crispin Wright.

Wright cares deeply about the topics of realism and antirealism, and sees insouciance about such issues as undesirable “quietism,” defined as the view that “significant metaphysical debate is impossible.”<sup>26</sup> James's and Dewey's pragmatism was, among other things, an attempt to shut off such debate—not by showing it to be impossible or senseless, but by

showing it to be pointless. So Wright's *Truth and Objectivity* is a good example of contemporary opposition to pragmatism. If the argument of that book is on the right track, then pragmatism is merely an unhappy attempt to evade questions that are absolutely central to philosophical reflection.

Like Davidson, Wright distrusts deflationism. But his reasons are very different. For Davidson, Tarski failed to show us how to detect in nature the pattern his truth theories for specific languages exhibit. But for Wright what Tarski failed to give us is a *norm*. Wright thinks our statement-making practices are regulated by two distinct norms: warranted assertibility and truth. These two are, Wright says, "distinct in the precise sense that although aiming at one is, necessarily, aiming at the other, success in the one aim need not be success in the other."<sup>27</sup> From Wright's point of view, the trouble with deflationism is not that it does not tell you how to work up a truth theory for a given natural language, but that it does not even mention your duty to attain the truth. It leaves you thinking that you have done enough if you have done all the justifying you can.

That, of course, is just what pragmatists want you to think. Here, it seems to me, Davidson can happily concur with the pragmatists. For, as I have already suggested, I see no way to fit the idea of truth as a goal of inquiry into Davidson's account of what we need to say about truth. So in order to widen still further the gulf that yawns between Davidson's quietism and the metaphysical activism urged by Wright, I shall stress the entanglement of Wright's claim that truth is a distinct norm with his unpragmatic and anti-Davidsonian attempt to keep the notions of "correspondence" and "representation" alive.

Wright says that "deflationism ... is committed to the idea that warranted assertibility is the *only* norm operating over assertoric discourse."<sup>28</sup> But, he says, even the deflationist has to admit that "while 'is T' and 'is warrantably assertible' are normatively coincident, satisfaction of the one norm need not entail satisfaction of the other."<sup>29</sup> So, Wright concludes, "deflationism reinflates." But this argument seems insufficient. The fact that beliefs can be justified without being true does not entail that two norms are being invoked. Analogously, the fact that an action can be fully justified to a given audience and still not be the right thing to do does not show that we have two duties—one to justify our actions to

each other and another to do the right thing. It merely shows that what can be justified to some audiences cannot be justified to others.

Wright, however, has a more detailed argument for his claim that “deflationism is an inherently unstable view.”<sup>30</sup> He takes the deflationist to say that the content of the truth predicate is “wholly fixed” by what he calls the Disquotationalist Schema:

“*P*” is true if and only if *P*.

Then he says that there is an “explanatory biconditional link effected by the Disquotational Schema between the claim that a sentence is *T* and its proper assertoric use.”<sup>31</sup> He defines a predicate as “(positively) descriptively normative” just in case “participants’ selection, endorsement, and so on of a move is as a matter of fact guided by whether or not they judge that move is *F*.”<sup>32</sup> This enables him to conclude that

‘*T*’ is descriptively normative in the sense that the practices of those for whom warranted assertibility is a descriptive norm are exactly as they would be if they consciously selected the assertoric moves which they were prepared to make or allow in the light of whether or not the sentences involved were *T*.

He sums up by saying that “any actual assertoric practice will be just as it would be if *T* were a self-conscious goal.”<sup>33</sup> Although the behavior of those selecting for warranted assertibility will be the same as that of those selecting for truth, Wright thinks that we can distinguish two selections by asking whether they are “as a matter of fact guided” by one consideration rather than another.

But is it enough for there to be a fact of such guidance that the agent thinks there is such a fact?<sup>34</sup> Consider an analogy: I am trying to decide whether to prosecute my father for impiety. In the course of doing so I sometimes describe myself as trying to do what I am justified in thinking pious and sometimes as trying to obey the will of the gods. Socrates has pointed out to me that although the two criteria are normatively coincident, satisfaction of the first criterion does not entail satisfaction of the second—for my community, the one that has given me my sense of what counts as satisfactory justification, may be out of touch with the gods. Still, my hope of satisfying both criteria persists.

An atheist, however, may tell me that I am “as a matter of fact” guided by only one norm and have only one self-conscious goal—that only one

process of conscious selection is at work in my decision making. Since there are no gods, he says, there is no such thing as their will, and I cannot, even if I want to, obey the norm of conformity to that will. But I, of course, shall rejoin that this line of thought is reductionist and that my belief in the gods is enough to enable me to attempt to obey this norm. What norms one obeys, after all, is a matter of what norms one thinks one is obeying.<sup>35</sup>

I do not think that Wright should be happy with this line of defense against the atheist. For the force of his term “as a matter of fact guided by” disappears once one grants that a belief in guidance is proof of guidance. An imaginative agent who proliferates goals, and thus lights by which to perform the self-conscious selection of moves, will soon have more guidance systems going than we can shake a stick at. He will, for example, be trying to hit every bull’s-eye he aims at, to win all the archery competitions, to become known as a superb archer, to become world archery champion, to please the goddess Diana, and to find a sympathetic defender in the councils of the gods. He will see all of these as prescriptively coincident—they all lead him to perform exactly the same actions—while acknowledging that achievement of the last two goals may not be extensionally coincident with achievement of the first four. For he has heard rumors that Diana has long since lost interest in archery and is now into karate.

Wright must either concede that a goal is “descriptively normative” for an action if the agent thinks it is, or else give us a further criterion for detecting *real* descriptive normativity. I am not sure what such a criterion could look like. But if he cannot specify one, he may have to admit that, just as “deflationism reinflates,” so atheism retheologizes.

My own view is that attaining divine favor was indeed a goal distinct from hitting the target for religious archers and that attaining truth as distinct from making justified statements *is* a goal for metaphysically active inquirers. We metaphysical quietists deplore the fact that most people in our culture can be incited to this sort of activity. They still, alas, take seriously such bad, unpragmatic questions as “subjective or objective?” “made or found?” “*ad nos* or *in se*?” “socially constructed or for real?” But just as religious archers can be (and to some extent have been) replaced by atheist archers, so we pragmatists hope our culture will

eventually replace itself with the culture that James and Dewey foresaw. In that culture, the question “Are you trying to attain truth as well as to form justified beliefs?” would be greeted with the same puzzlement with which “Are you seeking divine favor?” is greeted by atheist archers.<sup>36</sup>

I shall return to the topic of cultural change at the end of this essay, but first I want to direct attention to Wright’s motive for emphasizing the difference between deflationism, which does not recognize that truth is a distinct norm, and his own brand of minimalism, which does. Wright has two aims in his book. The first is to give deflationism its due by admitting that “truth is not intrinsically a metaphysically heavyweight notion.”<sup>37</sup> This puts Wright in a position to rebuff “error-theorists” like John Mackie, who think it a mistake to apply the word “true” to moral judgments. For, as Wright rightly says, “the minimalist proposal is conservative of our ordinary style of thought and talk about the comic, the revolting and the delightful, the good, and the valuable, which finds no solecism in the description of contents concerning such matters as ‘true.’”<sup>38</sup> Defeating philosophers like Mackie is Wright’s first aim, and Davidson and Dewey would both applaud this project.

His second aim, however, is to make clear that “we do not, in going minimalist about assertoric content and truth, set ourselves on a fast track to quietism about traditional philosophical controversy concerning realism and objectivity.”<sup>39</sup> Wright thinks that

talk of ‘representation of the facts’ is not just admissible phrasing, a harmless gloss on talk of truth, but incorporates a philosophically correct—as we might say, *seriously dyadic*—perspective on the truth predicate (at least for discourses where realism is appropriate).

His deflationist opponent, he goes on to say, will insist that such talk “is additional metaphysical theory, foisted onto phrases which, while characteristic of the idea of truth, can be saved by a deflationary account and merit no such metaphysical interpretation.”<sup>40</sup>

Dewey or Davidson could hardly have expressed his quietistic antipathy to the notions of correspondence and representation with better words than those that Wright here puts in the deflationist’s mouth. James’s and Dewey’s post-Darwinian attempt to naturalize our self-image by dissolving the traditional oppositions between mind and nature

and between subject and object, as well as Davidson's later assault on the scheme—content distinction, are both nicely epitomized in the claim that our perspective on the truth predicate should *not* be “seriously dyadic.”

One of the great merits of Wright's very dense and argument-packed book is that he sees the need to say more than Dummett does about the pragmatic cash value of the ideas of “realism,” “representation,” and “correspondence.” He sees that the logical terminology made current by Dummett—in his explications of “realism” with the aid of notions like bivalence and failure of excluded middle—does not adequately capture the motives for traditional debates. He notes that David Wiggins attempted to remedy this defect by suggesting that a tendency toward convergence is a sufficient criterion for the applicability of such notions. But Wright criticizes Wiggins's suggestion on the ground that the presence of such a tendency would, for example, make judgments about the comic representational if, for some accidental sociohistorical reason, there was steady convergence toward consensus on the comic.

Wright is surely right that the idea of representationality, and thus of realism, needs to be explicated with the help of a notion that is neither merely logical nor merely sociological. But his choice of a candidate for such an intermediate notion is very revealing. He says that what lies behind the intuitive association of representationality with convergence is “the Convergence/Representation Platitude,” namely:

If two devices each function to produce representations, then if conditions are suitable, and they function properly, they will produce divergent output if and only if presented with divergent input.<sup>41</sup>

This so-called platitude is supposed to flesh out the intuitive difference between the cognitive and the noncognitive, and thus between discourses (e.g., physics) for which realism is appropriate and others (e.g., argument about what's funny) for which it is not. Wright says that in the latter example “the base—the sense of humor—may blamelessly vary from person to person.” But when it comes to reporting on the colors and shapes of middle-sized pieces of dry goods, or to astronomical theory, we can blame people for not getting them right, not representing accurately, not living up to their cognitive responsibilities, not corresponding to reality.



One might think, however, that blamability itself might blamelessly vary for contingent sociohistorical reasons. Wright sees this point and grasps the nettle. Metaphysical questions, such as those about the cognitive status of a discourse, can, he says, be settled only *a priori*. He boldly offers the following definition:

A discourse exhibits Cognitive Command if and only if it is *a priori* that differences of opinion arising within it can be satisfactorily explained only in terms of 'divergent input'; that is, the disputant's working on the basis of different information (and hence guilty of ignorance or error, depending on the status of that information) or 'unsuitable conditions' (resulting in inattention or distraction and so in inferential error, or oversight of data and so on) or 'malfunction' (for example, prejudicial assessment of data, upwards or downwards, or dogma, or failing in other categories already listed).<sup>42</sup>

One might paraphrase this definition by saying that you are under Cognitive Command if you are functioning as a well-oiled representation machine. The picture Wright is using is the one used by all epistemologists who think of "prejudice" and "superstition" as sand in the wheels, the sort of foreign ingredient that causes malfunctions. Such philosophers share a picture of human beings as machines constructed (by God or Evolution) to, among other things, get things right. Pragmatists want our culture to get rid of that self image and to replace it with a picture of machines that continually adjust to each other's behavior, and to their environment, by developing novel kinds of behavior. These machines have no fixed program or function; they continually reprogram themselves so as to serve hitherto undreamt-of functions.

Wright's so-called platitude suggests that pragmatists should do to him what he did to Wiggins. We should say that representation drops out for the same reasons convergence did. When we drop both notions, what we are left with is their common cash value: the claim that it is a demarcating mark of the appropriateness of realism for a given discourse that a certain picture be applicable to that discourse: the picture of truth as the output of a well-functioning machine that incarnates an *a priori* knowable input-output function. Notice that it is not enough for Wright's purposes if we merely know *a priori* that some input-output function or other is at work and that failure of the machine to operate in accord with this function is a malfunction. That requirement will be uninterestingly satisfied by indefinitely many functions, and equally uninterestingly unsat-

ified by equally many others. What Wright requires is that we should know a priori which of these functions is the *right* one—that our knowledge of the *content* of the output (for example, the comic, the geometric, the valuable) should pick out a particular function.<sup>43</sup>

I shall return to this last point later, when I take up Wright's response to McDowell's argument for quietism. For the time being, however, I simply note that pragmatists, particularly after reading Kuhn, discard the terms "prejudice" and "dogma," as well as the idea that before the New Science came along, with its prejudice-detecting rationality and superstition-dissolving rigor, our cognitive machinery malfunctioned.<sup>44</sup> Pragmatists doubt that cognitivity amounts to more than historically contingent consensus about what shall count as proper justification for a belief. They see such consensus as what distinguishes what Kuhn calls "normal science" from what he calls "revolutionary science." Whereas Wright thinks that philosophers can look at the "content" of a discourse<sup>45</sup> and decide the a priori question of whether it is apt for Cognitive Command, pragmatists see the aptness or inaptness of Wright's "representation machine" terminology as up for historicosociological grabs—as much up for such grabs as the aptness or inaptness of religious language for describing the human situation.

Pragmatists think that Wright's "Consensus/Representation Platitude" can be made plausible only if one specifies that the two devices in question were machines for representing something *according to the same conventions*. For wildly different outputs can count as representations of the same input, depending on the purpose that the representational machinery serves. A videotape, an audiotape, and a typed transcript represent the same press conference. Anything, indeed, can count as a representation of anything, if there is enough antecedent agreement that it will count as such. More generally, representationality, and thus cognitivity, is something we can create, if not exactly at will, at least by agreement.

Content, pragmatists say on the basis of this argument, counts for vanishingly little in determining cognitivity, and de facto agreement on conventions for everything. That is why pragmatists think cognitivity a purely empirical, historicosociological notion. But if conventions of representation can vary as blamelessly as sense of humor—or, more to the

point, if the only relevant sort of blame is the sort that attaches to those who are insufficiently cooperative in achieving shared practical goals—then representationality, like convergence, is a broken reed. It is of no help in pinning down the nature of cognitivity or in offering a seriously dyadic account of truth.

How destructive to his overall program is this objection to Wright's putative platitude? I think the best way to find out is to turn to the only place in his book where Wright explicitly argues against quietism—his final chapter. For there he discusses a notion, "meaning," which is close kin to that of "convention of representation." The only argument in favor of quietism that he discusses is Wittgenstein's "rule-following argument."

Wright agrees with Gareth Evans that this argument is a "metaphysical wet blanket," to be tossed off if at all possible. For Wright, the *only* concession to the quietist that need be made is the one made in his first chapter: that truth and falsity can be had even where realism is out of the question (as it certainly is in the case of comedy and may be in the case of morals).<sup>46</sup> He considers two brands of wet Wittgensteinian blanket: McDowell's and Kripke's. For McDowell, Wright says, the moral non-cognitivist is "driven by a misbegotten construal of *ethical* fact and objectivity"; like her Platonist, moral realist opponent, she labors

under the misapprehension that anything worth regarding as moral cognitivism has to make out how the relevant subject matter is *there*, so to speak, for any enquiring agent, independently of an evaluative 'point of view'. Since, as Wittgenstein teaches us, no subject matter is ever 'there' in that kind of way, no disadvantageous comparison remains to be made. The appreciation of moral fact requires, to be sure, a moral point of view. But then, the appreciation of *any* fact requires a point of view.<sup>47</sup>

Wright rejects this attempt to undermine "realist/anti-realist debate in general."<sup>48</sup> He thinks that one great advantage of his notion of Cognitive Command is that it involves no "hyper-objectified conception of fact" of the sort Wittgenstein and McDowell criticize. For "the question whether it is a priori that differences of opinion formulable within a particular discourse will, prescinding from vagueness, always involve something worth regarding as a cognitive shortcoming" is to be settled "by reference to what we conceive as the range of possible sources of such differences."<sup>49</sup>

This, however, is not a sufficient reply to McDowell. For to have a conception of the range of possible sources of such differences, we first need to specify an input-output function; without that, we will not be able to distinguish the smooth functioning of a representation machine from its malfunction. But many input-output functions will describe the machine, and not all these functions will range over the same inputs. There are many ways to classify the flux of causal interactions in which the statement maker is involved, and each will offer a new candidate for “input.” The problem of whether Wittgenstein has in fact shown that the relevant subject matter is never “there” in the relevant sense is the problem of whether there is a way to isolate input without reference to what Wright calls “an evaluative point of view.” This is the sixty-four-dollar question: whether we can (as Dewey and Davidson insist we cannot) separate out “the world’s” contribution to the judgment-forming process from our own.

Wright has no doubts about the existence of isolable truth makers. At one point, for example, he says that

the world’s making such statements [those that are what he calls “super-assertible”] likely is something conceptually quite independent of our standards of appraisal—something, as it were, which is wholly between the statement and its truth maker, and on which we impinge only in an (at most) detective role.<sup>50</sup>

But how are we supposed to separate out these truth makers from the flux of causal interactions in which the statement maker is involved? One of Davidson’s reasons for having no truck with the idea of “truth makers”<sup>51</sup> is his hunch that only completely artificial objects called “facts”—what Strawson sneeringly called “sentence-shaped objects”—can meet Wright’s needs. The problem is not with funny, Platonic, “hyper-objectivized” facts, but with *any* sentence-shaped nonsense, any putatively (in McDowell’s words) “nonconceptualized configurations of things in themselves.” Insofar as they are nonconceptualized, they are not isolable as input. But insofar as they are conceptualized, they have been tailored to the needs of a *particular* input-output function, a *particular* convention of representation.<sup>52</sup>

That any causal transaction can exemplify many different input-output functions was, of course, Wittgenstein’s point when he remarked that all

my previous additions could be seen as satisfying indefinitely many different rules for the use of “plus.” But it is only when Wright turns from McDowell’s Wittgenstein to Kripke’s that he takes up this sort of difficulty explicitly. In discussing Kripke’s, he considers the possibility that “the thesis that there are no ‘facts of the matter’ as far as rules and meanings are concerned . . . must necessarily inflate . . . into a *global* irrationalism: the thesis that there are no facts of the matter anywhere.” For

if there are no substantial facts about what sentences say, there are no substantial facts about whether or not they are true. Thus, irrationalism about meaning must enjoin an irrationalism about truth, wherever the notion is applied. And irrationalism about truth, wherever the notion is applied, is irrationalism about all assertoric discourse.<sup>53</sup>

On this account of what Wittgenstein was up to, the problem is not, as with McDowell’s Wittgenstein, that the indefinite plurality of rules (or conventions of representations, or input-output functions) makes it impossible to draw an interesting representational-nonrepresentational line between discourses, but that we have (by some criterion of non-representationality that remains obscure) discovered that there is no such thing as getting meanings right, no such thing as representing meanings accurately.

Wright has an answer to this suggestion, one that I found very hard to follow and will not try to summarize.<sup>54</sup> But it seems clear that this suggestion is not the interesting one. For the interesting question about quietism, the one to which Wright’s final chapter is devoted, is whether the whole terminology of “getting right” and “representing accurately” is a useful way of separating off discourses from one another. This question, raised by McDowell’s Wittgenstein, is begged by Kripke’s.

Wright seems to recognize this point, for in the penultimate paragraph of his book, he grants that there is a “residual concern” to which he has not spoken. The following is his final formulation of this concern:

[W]hether, even if the key distinctions [between representational, cognitive, substantive truth, and the other, merely minimalist sort of truth] can be formulated in ways that allow the status of a discourse to be determined independently of the rule-following dialectic, their serviceability as vehicles for the expression of realist intuition may not be so severely compromised by a proper understanding of that dialectic that there is no longer any *point* to the taxonomy which they might enable us to construct?<sup>55</sup>

Raising this doubt—a doubt about whether there was a point to the book we have just finished reading—on the book's last page seems to me a very honest, and rather brave, thing to do.

Wright's response to this doubt, in his final paragraph, is that though there may be a case to be made for the view that there is no point, his book has helped set the terms for debating any such case by giving us "a more pluralistic and fine-grained conception of the realist/anti-realist debates than has been commonplace."<sup>56</sup> It has indeed given us such a conception, but the increased fineness of grain may not serve the purpose Wright suggests. For what looks like desirable fineness of grain to Wright looks like the pointless addition of further epicycles to his pragmatist opponents.

Wright's suggestion—which, though I have scanted it for my purposes in this essay, is at the heart of his book—is that there are different truth predicates for different discourses. He argues that we should use a minimalist, thin truth predicate in discourse about the comic, and various thicker alternatives (such as Cognitive Command), correlated with various *a priori* determinable relations between other discourses and the rest of the world, for other discourses.<sup>57</sup> But of course for pragmatists, what Wright thinks of as permanent *a priori* determinable relations are just local and transitory *historicosociological* differences between patterns of justification and blame. These differences—subpatterns within the single overall pattern justification makes—should not, pragmatists think, be imported into the concept of truth. To do so is to do what Davidson calls "humanizing truth by making it basically epistemic."<sup>58</sup>

Much of what I have said can be summed up in the claim that the central issue between Wright's metaphysical activism and Davidson's quietism concerns the point of inquiry. For Wright truth, considered as a desirable noncausal relation between language and nonlanguage, is a goal of such inquiry (if only in those areas of culture, such as physical science, for which "realism" is thought appropriate). For Davidsonians, on the other hand, the most consistent position is to hold that

(a) the arguments from the indefinite plurality of ways of going on/input-output functions/conventions of representations leave no room for any such desirable noncausal relation,

- (b) so there is no reason to think that even an infinite amount of justification would get us closer to such a relation,
- (c) so there is nothing that can plausibly be described as a *goal* of inquiry, although the desire for further justification, of course, serves as a *motive* of inquiry.

If Dewey and Davidson were asked, "What is the goal of inquiry?" the best either could say would be that it has many different goals, none of which have any metaphysical presuppositions—for example, getting what we want, the improvement of man's estate, convincing as many audiences as possible, solving as many problems as possible. *Only* if we concede to Wright that "truth" is the name of a distinct norm will metaphysical activism seem desirable. For Dewey and Davidson, that is an excellent reason not to view it as such a norm.

Some Davidsonians might see no reason why they too should not say, ringingly, robustly, and commonsensically, that the goal of inquiry is *truth*. But they cannot say this without misleading the public. For when they go on to add that they are, of course, not saying that the goal of inquiry is correspondence to the intrinsic nature of things, the common sense of the vulgar will feel betrayed.<sup>59</sup> For "truth" sounds like the name of a goal only if it is thought to name a *fixed* goal—that is, if progress toward truth is explicated by reference to a metaphysical picture, that of getting closer to what Bernard Williams calls "what is there *anyway*." Without that picture, to say that truth is our goal is merely to say something like: we hope to justify our belief to as many and as large audiences as possible. But to say that is to offer only an ever-retreating goal, one that fades forever and forever when we move. It is not what common sense would call a goal. For it is neither something we might realize we had reached, nor something to which we might get closer.

We pragmatists think that philosophers who view the defense of "our realistic intuitions" as an important cultural or moral imperative are held captive by the picture of getting closer to a fixed goal. As an initial step in breaking free of this picture, we suggest following Davidson in abandoning what he calls "standard ideas of language mastery." Then one will think of such mastery as involving "no learnable common core of consistent behavior, no shared grammar or rules, no portable inter-

preting machine set to grind out the meaning of an arbitrary utterance.”<sup>60</sup> Dropping these standard ideas makes it very difficult to take seriously the idea of human beings as portable representing machines that incorporate a priori knowable input-output functions.

The idea of such a machine lies behind both Wright’s notion of Cognitive Command and his Kripkean suggestion that language, meaning, truth, and knowledge might all collapse together if, *horribile dictu*, it should turn out that there is no fact of the matter about what we have been meaning by “addition.” But the skepticism described by Kripke’s Wittgenstein holds no terrors for those who follow Davidson in abandoning the whole idea of “rules of language.” Analogously, skepticism about an a priori recognizable attribute of discourses called cognitivity or representationality, and about the utility of the notions of cognitivity and representationality, holds no terrors for those who, like Bacon, Dewey, and Kuhn, see artisans and natural scientists as doing the same kind of thing: employing whatever propositional or nonpropositional tools they think may help with the problems currently before them.<sup>61</sup>

If, as good Darwinians, we want to introduce as few discontinuities as possible into the story of how we got from the apes to the Enlightenment, we shall reject the idea that Nature has settled on a single input-output function that, incarnated in each member of our species, enables us to represent our environment accurately. For that idea requires that Nature herself has divided up the causal swirl surrounding these organisms into discrete inputs and has adopted a particular input-output function as distinctively hers—a function whose detection enables us to offer justification according to Nature’s own criteria (or, as Wright would say, Commands) rather than to those of transitory and local audiences. So, for Darwinians, there is an obvious advantage in *not* dividing the activities of these organisms into the cognitive, representational ones and the others. This means that there is an obvious advantage in dropping the idea of a distinct goal or norm called “truth”—the goal of scientific inquiry, but not, for example, of carpentry. On a Deweyan view, the difference between the carpenter and the scientist is simply the difference between a workman who justifies his actions mainly by reference to the movements of matter and one who justifies his mainly by reference to the behavior of his colleagues.



In previous essays—in particular one called “Science as Solidarity”<sup>62</sup>—I have urged that the romance and the idealistic hopes that have traditionally been elaborated in a rhetoric of “the pursuit of objective truth” can be equally well elaborated in a rhetoric of social solidarity—a rhetoric that romanticizes the pursuit of intersubjective, unforced agreement among larger and larger groups of interlocutors. But I agree with those who insist that the former rhetoric is that of contemporary common sense. So I think that pragmatism should not claim to be a common-sensical philosophy. Nor should it appeal, as David Lewis suggests metaphysics must appeal, to intuition as final arbiter.

If contemporary intuitions are to decide the matter, “realism” and representationalism will always win, and the pragmatists’ quietism will seem intellectually irresponsible. So pragmatists should not submit to their judgment. Instead, they should see themselves as working at the interface between the common sense of their community, a common sense much influenced by Greek metaphysics and by monotheism, and the startlingly counterintuitive self-image sketched by Darwin, and partially filled in by Dewey. They should see themselves as involved in a long-term attempt to change the rhetoric, the common sense, and the self-image of their community.

The pragmatist who says, “The difference between justification and truth makes no difference, except for the reminder that justification to one audience is not justification to another”—the claim I put in her mouth at the beginning of this essay—has not yet said enough. For there is another difference: justification does not call for metaphysical activism but truth, as understood by contemporary, representationalist common sense, does. The pragmatist regrets the prevalence of this representationalist picture and of the “realist” intuitions that go with it, but she cannot get rid of these unfortunate cultural facts by more refined analyses of contemporary common sense. She cannot appeal to neutral premises or to widely shared beliefs.

She is in the same situation as are atheists in overwhelmingly religious cultures. Such people can only hope to trace the outlines of what Shelley calls “the gigantic shadows which futurity casts upon the present.” They foresee a time when the notions of Divine Will and of Cognitive Command will, for similar reasons, have been replaced by that of a Free

Consensus of Inquirers. But, in the meantime, the pragmatist who urges our culture to abandon metaphysical activism cannot argue that such activism is inconsistent with a mass of our other beliefs, any more than ancient Greek atheists could say that sacrificing to the Olympians was inconsistent with a mass of other Greek beliefs. All the pragmatist can do is the sort of thing they did: she can point to the seeming futility of metaphysical activity, as they pointed to the seeming futility of religious activity.

In the end, we pragmatists have no real arguments against the intuitions to which books like Wright's *Truth and Objectivity* appeal. All we have are rhetorical questions like: Are all those epicycles really worth the trouble? What good do the intuitions you painstakingly salvage do us? What practical difference do they make?<sup>63</sup> But such rhetorical questions have been instruments of sociocultural change in the past, and may be again.

#### Notes

1. Of course, when the question is not about deciding what to believe now, but about explaining what has happened, the distinction between justification and truth is useful: we often explain our failures by saying, "I was quite justified in believing that, but unfortunately it was not true." But though useful, it is not essential. We can explain our failure equally well by saying "What I thought would happen did not," and in many other ways.
2. However, what there is to be said about justification is local rather than global: quite different, unconnected things have to be said about justification in, for example, mathematics, jurisprudence, and astrology. So philosophers are hardly the people to say it. This point chimes with Michael Williams's argument (in his *Unnatural Doubts: Epistemological Realism and the Basis of Scepticism* [Oxford: Blackwell, 1991]) that "knowledge" is neither the name of a natural kind nor the topic of useful global theorizing. I am indebted to Williams for the realization that the Cartesian notion of a natural, ahistorical, and transcultural "order of reasons" is essential to Descartes's "dreaming" argument, and more generally to both epistemological skepticism and the feasibility of epistemology as a discipline.
3. Not all. Some idealists argued that all truths are true by virtue of their correspondence to a single object (the Absolute), thereby eviscerating the idea of correspondence.
4. Of course, a host of contemporary philosophers (notably Ruth Millikan, David Papineau, and Fred Dretske) have retained the notion of "inner representation"

and interpreted it biologically, as a matter of the evolutionarily designed ability of an organism to respond differentially to different stimuli. In contrast, followers of Wilfred Sellars (such as George Pitcher, David Armstrong, Daniel Dennett, and myself) lump the neurological arrangements that make possible such differential responses to differential stimuli together with the internal states of (for example) thermostats. We treat perceptions as dispositions to acquire beliefs and desires rather than as “experiences” or “raw feels,” and hence we disagree with Thomas Nagel that there is “something it is like” to have a perception. I see the Sellarsian strategy we employ as an example of the pragmatist habit of refusing to recognize the existence of troublemaking entities. This habit strikes nonpragmatists like Nagel as a refusal to face up to the facts.

As I suggest at the end of this essay, we pragmatists too want to be faithful to Darwin. But we think that the Millikan-Papineau-Dretske revivification of the notion of “representation” is an insufficiently radical way of appropriating Darwin’s insight. These philosophers want to reconcile Darwin with Descartes’s and Locke’s “way of ideas.” In contrast, we want to follow up on Dewey’s suggestion that Darwin has made Descartes and Locke obsolete.

5. *Pragmatism and The Meaning of Truth* (Cambridge, Mass.: Harvard University Press, 1975), 106.

6. *Ibid.*, 42. James also, unfortunately, said a lot of other, conflicting things about truth—such as that it consists in some kind of agreement between ideas and reality. In “Dewey Between Hegel and Darwin” (reprinted in *Truth and Progress: Philosophical Papers*, vol. 3 [Cambridge: Cambridge University Press, 1998]), I argue that Dewey was wise to avoid saying the latter sort of thing and to eschew analyses or definitions of “truth” or of “true.”

7. Two recent books show how this suggestion can be worked out in detail: Barry Allen’s *Truth in Philosophy* (Cambridge, Mass.: Harvard University Press, 1993) and Robert Brandom’s *Making It Explicit* (Cambridge, Mass.: Harvard University Press, 1994).

8. For an account of this strategy, see Hilary Putnam’s “Does the Disquotational Theory Solve All Problems?” in his *Words and Life* (Cambridge, Mass.: Harvard University Press, 1994), 264–78. Putnam there criticizes two philosophers whom he construes as disquotationalists—Paul Horwich and Michael Williams—for remaining in the grip of a “positivistic picture” and for being closet reductionists. This is a criticism he has often made of me (see, e.g., “The Question of Realism,” 295–312, in the same volume). On Putnam’s view, all three of us ignore the need to admit the existence of genuine “directedness” and “intentionality.” I am not sure whether Putnam would make the same criticism of Davidson.

9. Richard Rorty, “Pragmatism, Davidson and Truth,” in *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, ed. Ernest LePore (Oxford: Blackwell, 1986), 333–68. This article is reprinted in my *Objectivity, Relativism, and Truth* (Cambridge University Press, 1991).

10. Donald Davidson, “The Structure and Content of Truth,” *Journal of Philosophy* 87, no. 6 (1990), 279–328. This article comprises Davidson’s Dewey Lectures.

11. Ibid., 287.

12. Ibid.

13. Ibid., 295.

14. Ibid.

15. Ibid., 322.

16. Ibid., 323.

17. Ibid., 320.

18. A good statement of the view that you *can* separate these is Wright's description of metaphysical realism, as asserting the possibility that "despite the apparent cognitive richness of our lives, we are somehow so situated as not to be enabled to arrive at the concepts which fundamentally depict the character of the real world and the nature of our interaction with it" ("Putnam's Proof That We Are Not Brains in a Vat," in *Reading Putnam*, ed. Peter Clark and Bob Hale [Oxford: Blackwell, 1994], 238). Assuming that "fundamentally depict the character of ..." means "are required to tell the truth about," then Davidson is committed to saying that this situation cannot arise: there can never be what Wright calls "a thought whose truth would make a mockery of humankind and its place in nature" (ibid., 240). The worst that can happen is that people whose language we are quite capable of learning (the Galactics, say) might offer us some astonishingly impressive substitutes for our present beliefs about selected special topics (e.g., the microstructure of matter or how to achieve world peace).

19. Davidson, "Structure," 309. Some commentators on Davidson have taken a truth condition to be a nonlinguistic state of affairs, a fact rather than a sentence in the truth theorist's language—despite Davidson's polemic against the notion of "fact" in "True to the Facts," in *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press, 1984), 36–54.

20. Davidson, "Structure," 309.

21. Michael Williams in *Unnatural Doubts* suggests that an inability to "account for the truth-conduciveness of justification" will lead to skepticism (231). My view, and the one I am attributing to Davidson, is that what leads to skepticism is the initial assumption of truth-conduciveness rather than the failure of attempts to back up this assumption. So I deny Williams's claim that "it is surely an essential feature of epistemic justification that justifying a belief makes it more likely to be true" (229). I enlarge on this denial in "Sind Aussagen universelle Geltungsansprüche?" (*Deutsche Zeitschrift für Philosophie* 42, no. 6 [1994], 975–88), a criticism of Habermas's and Apel's views on truth.

22. For his repudiation of the notion of "representation," see Davidson's "The Myth of the Subjective," in *Relativism: Interpretation and Confrontation*, ed. Michael Krausz (Notre Dame, Ind.: Notre Dame University Press, 1989), 165–6.

23. But I may be missing something here, and my blind spot may conceal a real and important disagreement between Davidson's views and my version of pragmatism. For in "Structure," Davidson says that "[s]ince the concept of truth is

central to the theory [i.e., to an empirical theory that entails T-sentences], we are justified in saying that truth is a crucially important explanatory concept" (313). It does not look particularly central to me. As I see it, what Davidson calls a "theory of truth" could equally well be called "a theory of complex behavior" or "a theory of justificatory behavior." Granted that the production of the sort of biconditionals Tarski called "T-sentences" is the whole point of the theory, I am not sure why the production of these sentences illustrates the centrality, or the crucial importance, of the concept of truth.

I am quite willing to withdraw my 1986 claim that "true" has no explanatory use, which was a misleading way of putting the point that "It's true!" is not a helpful explanation of why science works or of why you should share one of my beliefs. But although the sort of theory to which Davidson thinks "the concept of truth" central is indeed explanatory, it seems to me somewhat awkward and unnecessary to pick out a given concept that is explicated by reference to such theories and say that *it* has a crucial explanatory role. Avoiding such favoritism would be more congruent with Davidson's fundamental point that a theory of truth is automatically a theory of meaning and of rationality—as well as with his doctrine that every intensional concept is intertwined with every other such concept.

Another way of locating the point at which Davidson and I may differ is that he thinks it significant that we use the same word to designate what is preserved by valid inference as we use to caution people that beliefs justified to us may not be justified to other, better audiences. As far as I can see, there is no deep reason why "true" is used to do both of these jobs, why one of the words that we use to describe the pattern of behavior necessarily exhibited by language users (logical inference) should also be one of the words we use to caution people that they may be believing something that better-advised people would not believe. So I see no reason to look behind both uses for some feature of the *meaning* of "true" which makes that word suitable for both assignments. If I could see such a reason, I might be in a better position to appreciate what Davidson means by the "centrality" of the concept and to see why he speaks of himself as "filling in the content" of this concept.

24. Davidson, "Structure," 319.

25. *Ibid.*, 325.

26. Crispin Wright, *Truth and Objectivity* (Cambridge, Mass.: Harvard University Press, 1992), 202.

27. *Ibid.*, 19.

28. *Ibid.*, 21.

29. *Ibid.*, 23.

30. *Ibid.*, 34.

31. *Ibid.*, 17.

32. *Ibid.*, 16.

33. *Ibid.*, 17.

34. Wright identifies the claim that to possess truth is “to meet a normative constraint distinct from assertoric warrant” with the claim that “truth is a genuine property” (ibid., 35). I avoid the issue of whether truth is a property—an issue that seems to me to boil down eventually, just as Wright says, to the question “one norm or two?” I agree with what Davidson says about this issue (“Structure,” 285).

35. This line of argument is often employed against, for example, a Hobbesian reductionist who says that the actions I think are motivated by my desire to be a good citizen are really motivated by my fear of sanctions. Hobbes’s and Thrasymachus’s strong point is that a causal explanation of my action that does not refer to good citizenship may be as useful as one that does. Their opponent’s strong point is that the need for causal explanation is not our only motive for attributing motives.

36. These last six paragraphs are heavily indebted to Bjorn Ramberg and Barry Smith. They replace a section of an earlier version of this essay, a version read and discussed by Smith and Ramberg. Ramberg kindly conveyed Smith’s (well-taken) criticisms of that version to me and suggested ways to avoid these criticisms—suggestions I have gratefully adopted.

37. Wright, *Truth and Objectivity*, 72.

38. Ibid., 75.

39. Ibid., 86.

40. Ibid., 83.

41. Ibid., 91.

42. Ibid., 93.

43. I cannot figure out how somebody who invokes a priori knowledge as blithely as Wright does can say, equally blithely, that “apriority generally is an artifact of description” (ibid., 129).

44. Wright has read Kuhn too, of course, and discusses “theory-ladenness” in some detail. But the upshot of his discussion is rather disappointing: “[T]he hope must be either that we can yet win through to some purified notion of an observation statement, one that does not involve ‘theory-ladenness’ of the sort which is giving the trouble, or—more likely—that the Cognitive Command constraint can and must be refined in some way while remaining faithful to its motivation in the idea of representational function. I have no easy solution to suggest” (167–8).

This passage is typical of Wright’s hope to smooth over the anomalies that arise from attempts to make explicit the presuppositions of traditional, intuitive distinctions. Pragmatists rejoice in no longer needing to invoke those distinctions or to have those intuitions. So what looks like undesirable quietism to Wright looks like vigorous philosophical progress to them. This is the same sort of dialectical standoff that obtained between Leibniz and Newton. Newton shrugged off, quietistically, many traditional Aristotelian problems. Leibniz insisted that such shrugs were symptoms of intellectual irresponsibility and that metaphysical, as well as physical, explanations were required.

45. "If our interest is in the question whether comic discourse, *by virtue of its very content*, is fitted to express the products of a seriously representational mode of function, then any constraint designed to capture that idea must, it seems, be so formulated that satisfying it requires the possibility of a priori knowledge that the relevant conditions are met" (ibid., 94). In Wright's usage, it seems to me, the concepts "content of a discourse" and "a priori knowledge about that discourse" are mutually definable. For the only way one would know whether one had zeroed in on the content of a discourse, as opposed to the mechanisms of its production, would be to figure out what could be known about that discourse a priori. For Quinean holists like Davidson, of course, these mutually definable concepts are equally dubious.

46. "... quietism makes at least one important contribution, viz., the insight that it is a metaphysical hypostasis of notions like truth and assertion to write their applicability within a discourse into the substance of a realist view about its subject matter" (ibid., 204).

47. Ibid., 207.

48. Ibid., 208.

49. Ibid., 208.

50. Ibid., 77. Elsewhere Wright speaks of "the kind of state of affairs conferring truth on *P*" (117). The homiletic tone of "between the statement and its truth-maker" recurs when Wright says that "where we deal in a purely cognitive way with objective matters, the opinions which we form are in no sense optional or variable as a function of permissible idiosyncrasy, but are *commanded* of us" (146). Contrast the *libertinisme erudit* implicit in the concluding words of Davidson's "Structure": "... truth thus rests in the end on belief and, even more ultimately, on the affective attitudes" (326).

51. See Davidson, *Inquiries into Truth*, 194, for his rejection of the idea of "truth makers." See also, in the same volume, his 1969 essay "True to the Facts," which contrasts the Tarskian notion of a sentence being satisfied by objects that can be individuated without the use of the sentence with that of "a sentence being made true by a fact." Since writing that essay, however, Davidson has dropped the claim that the former notion gives us any sort of correspondence account of truth.

52. Wright seems to be speaking to this issue when, in a pregnant, compressed, and baffling footnote, he says that "it seems just plain obvious that the reaction-dependence of rules, the ceaseless involvement of our sub-cognitive natures in our step-by-step appreciation of the requirements of rules which Wittgenstein emphasizes, cannot be at ease with the mythology of the epistemically transparent yet fully substantial propositional object" (226). He suggests that the moral to be drawn from Wittgenstein may be that "something irreducibly *human and sub-cognitive* actively contributes to our engagement with any issue at all—a contribution ... presumed shared among thinkers who engage the issue in question" (227). If that is what Wittgenstein told us, he chose a remarkably roundabout way of saying that we can presume that our interlocutors' bodies respond to

the environment pretty much as ours do. One difference between Wright and Davidson is that Davidson would, I think, see no point in distinguishing between a cognitive nature or level and a noncognitive nature or level, for the same reasons he sees none in distinguishing between scheme and content, or between subject and object, or between “knowing a language and knowing our way around the world generally” (“A Nice Derangement of Epitaphs,” in *Truth and Interpretation*, ed. LePore, 445–6). From his point of view, such distinctions hypostatize two descriptions, one in propositional and one in nonpropositional terms, of the same events.

I confess, however, that Davidson’s attachment to the doctrine of the indeterminacy of translation, and his related insistence that there is a philosophically interesting difference between the intentional and the nonintentional, suggest that he qualifies the thoroughgoing antidualism I am attributing to him here. I discuss this attachment in “Davidson’s Mental-Physical Distinction,” forthcoming in *The Philosophy of Donald Davidson*, ed. Lewis Hahn, The Library of Living Philosophers (La Salle, Ill.: Open Court).

53. Wright, *Truth and Objectivity*, 211.

54. See *ibid.*, 227.

55. *Ibid.*, 229–30.

56. *Ibid.*, 230.

57. “There are a *variety* of features that may be possessed by minimally truth-apt discourses, any of which may contribute in some measure towards clarifying and substantiating realist preconceptions about it. . . . A basic realism about a discourse (of course, the epithets “realism” and “anti-realism” come to seem less and less happy from a pluralistic perspective) would be the view that it is qualified by *no* interesting feature serving to give point to an intuitive realism about it—that it deploys minimally truth-apt contents, and that’s the whole of the matter” (*ibid.*, 141–2).

58. Davidson, “Structure,” 298.

59. For a good example of the outrage that results from such betrayal see John Searle, “Rationality and Realism: What Is at Stake?” *Daedalus* 122, no. 4 (1993), 55–83. Searle believes that there are ways of getting around the traditional problems with the notion that truth is accuracy of representation (65–6) and that those of us (he mentions Kuhn and Derrida, as well as myself) who think these problems insoluble are, by departing from what he calls the “Western Rationalistic Tradition,” endangering the universities. I reply to Searle’s article in “John Searle on Realism and Relativism” (*Truth and Progress*, chap. 3).

60. Davidson, “A Nice Derangement,” 445. Dropping these ideas also makes it very difficult to get excited about Wittgenstein’s rule-following argument. For freedom from these ideas permits one to see it as simply a version (adapted to the needs of those who still take the notion of “rules of language” seriously) of a generic argument against the existence of any relation that is both natural (i.e., not simply a product of contingent human practices) and noncausal. That is



the sort of relation which representationalists are constantly invoking: For the Sellarsian version of this argument, see Brandom, *Making It Explicit*, and my “Robert Brandom on Social Practices and Representations,” in *Truth and Progress*, chap. 6.

61. Kuhn summed up his claim that science should be thought of as problem solving by saying, “[W]hether or not individual practitioners are aware of it, they are trained to and rewarded for solving intricate puzzles—be they instrumental, theoretical, logical or mathematical—at the interface between their phenomenal world and their community’s beliefs about it” (“Afterwords,” in *World Changes: Thomas Kuhn and the Nature of Science*, ed. Paul Horwich [Cambridge, Mass.: MIT Press, 1993], 338). Like Dewey, Kuhn thought that a historicosociological account of the origin of these interfaces and these puzzles can replace a metaphysical account of the nature of representation. “I aim,” he says, “to deny all meaning to claims that successive scientific beliefs become more and more probable or better and better approximations to the truth and simultaneously to suggest that the subject of truth claims cannot be a relation between beliefs and a putatively mind-independent or ‘external’ world” (ibid., 330).

62. Included in my *Objectivity, Relativism, and Truth*.

63. Putnam thinks we have more, namely demonstrations of the incoherence of nonpragmatic positions. In *Words and Life* (Cambridge, Mass.: Harvard University Press, 1994) he explains “incoherence” as the fact that “attempts at a clear formulation of the [metaphysical realist] position never succeed—because there is no real content there to be captured” (303). I think that clarity is a matter of familiarity rather than a property whose presence or absence can be demonstrated, and that Wright, Bernard Williams, and others find clear as the noonday sun what Putnam finds irremediably unclear. So I prefer to talk of lack of convenience rather than lack of clarity. James Conant discusses the metaphysical issue between Putnam and myself in his introduction to *Words and Life* (xxx–xxxi).

# IV

---

## Phenomenological and Postmodernist Conceptions



---

# Phenomenological and Postmodernist Conceptions: Introduction

## Heidegger: Truth as Disclosure

Martin Heidegger asks after the essence of truth and decides that it is freedom. Radical and mysterious, Heidegger's theory of truth directly challenges many of the assumptions made by the other contributors to this volume.

Heidegger's essay begins innocuously enough. He notes that the "usual" concept of truth is correspondence between thought and reality: "A statement is true if what it means and says is in accordance with the matter about which the statement is made." Heidegger rejects this definition, not because it is incorrect, but because it is superficial and uninformative. As other writers in this volume also note, it is difficult to know what to make of the correspondence theory of truth unless one understands what "correspondence" means. In Heidegger's view, one thing it cannot mean is "copying" or "similarity." If similarity were the key to truth, then statements would be completely true only if they became the things they were about, and hence were no longer statements at all. As they are, statements just aren't very similar to things. Instead, Heidegger sees propositional truth as presupposing a more "primordial" relation of accordance between humanity and beings in the world: what he calls variously "openness" or "unconcealedness." Truth is "disclosure of beings through which an openness essentially unfolds." Roughly, to speak truly is to uncover beings as they are.

By the "essence" of truth, Heidegger is interested in what makes truth possible, in its ground. As Heidegger notes, the Western philosophical tradition typically assumes that accordance of a statement with reality is

possible only because the world itself is intelligible or inherently structured in a unique way. Historically, this structure has been thought to be the will of God, and truth is possible because proposition and matter “are fitted to each other on the basis of the unity of the divine plan of creation.” But even if we drop the idea of divine creation, the assumption usually remains that an absolute “world structure” grounds the possibility of objective truth. Our thoughts are true when they conform to that structure. Heidegger’s radical alternative is that what makes it possible for our statements to be true is not the inherent structure of a divinely created world but human comportment. Crudely put, it is our way of being in the world that makes truth and falsity possible. He calls this way of being “freedom” and argues that it consists in “letting beings be.”

Heidegger’s theory challenges some of the more established philosophical assumptions about truth. First and most important, it challenges the idea that truth is a static, binary relation between a subject’s representation of an object and that object itself. In contrast, Heidegger views a true assertion as an act of directly presenting or disclosing beings. Thus truth is neither correspondence nor coherence, in the robust sense of either term, but the product of an activity that presents the world directly. Second, this suggests that truth depends on humanity in some sense: “There is truth—unveiling and unveiledness—only when and as long as Dasein exists” (1982, 219). It is difficult, perhaps, to know exactly what Heidegger meant by this claim, but part of his meaning here is simply that without human thinkers there would be no true thoughts. But Heidegger is also concerned to emphasize that it is only against the background of human interests and needs that parts of the world become possible objects of knowledge.

This raises the question as to what extent Heidegger was committed to a type of idealism. At no point does he suggest, as did the classical idealists, that the world is composed of ideas or is entirely mental in some way. But he does seem to think that many aspects of the world are as they are because of human thought. Ultimately, much of the matter may hang on what Heidegger means by saying that truth consists in “letting beings be.” Given his insistence that “letting beings be” involves human engagement with the world, the phrase apparently does not mean letting beings be as they are determined to be by the inherent structure of the

world. But if not, then in what sense are we “letting” as opposed to *making* beings as they are?

### **Foucault: Truth as Power**

As Richard Rorty has noted, “postmodern” is an overused term that frequently causes more trouble than it is worth. But if we restrict ourselves to Lyotard’s sense, namely as “distrust of metanarratives,” then Foucault’s work on knowledge and truth is paradigmatically postmodern. A “metanarrative” is a transcendental theory or framework that can be used to judge or evaluate all other theories and frameworks, and a good example would be any one of the traditional theories of truth found in this volume. Foucault distrusts these traditional theories because he sees them as missing the most fundamental issue about truth, namely, its relation to political power.

In the selection included here (originally part of his written response to interview questions) Foucault suggests that truth “be understood as a system of ordered procedures for the production of, regulation, distribution, circulation and operation of statements.” This system is “linked in a circular relation with systems of power which produce and sustain it,” such as the scientific community, the government, and the media. Thus, in Foucault’s view, truth is by nature political.

On one reading, Foucault is advocating a form of skepticism about truth, and indeed many of his readers have taken him in this way (see, e.g., Allen 1993). On this interpretation, Foucault’s point is that there simply are no objectively true statements in the usual sense; there are only statements that “pass for true” in a particular community at a particular time (Allen 1993). And what passes for true is determined by the hegemonic systems of power. As Charles Taylor has remarked (1986, 93), Foucault’s seemingly Nietzschean take on truth totally subordinates, or reduces, truth to power.

If so, then Foucault’s view faces at least two problems. First, if truth is nothing more than what passes for truth and what passes for truth is constituted by systems of power, then as those systems change, so does the truth. It follows that in the American South of the 1960s and 1970s, African-Americans really were morally and intellectually inferior to

whites because that was the view of the white political-power structure at the time. But this is anti-intuitive: surely such racist views were false then *and* are false now. Second, a radical skepticism about truth would also seem to pull the rug out from under the feet of any attempt to criticize the political systems of power in one's own culture. For according to the view in question, a critic's statements can be true only if they are part of a system of power. Now either (a) these statements are part of the system they attempt to criticize, or (b) they are part of another system. Neither alternative seems to allow for meaningful social criticism: (a) implies that one must always be in agreement with the system one is trying to criticize, while (b) boils down to simply opposing one system of dominance with another.

Linda Alcoff (1995) has suggested an alternative reading of Foucault's understanding of truth. In Alcoff's view, Foucault's remark that truth is a "system of ordered procedures" should be seen as a version of the coherence theory of truth, one that is consistent with "the possibility that most commonsense and simple perceptual beliefs may remain constant through discursive changes" (Alcoff 1995, 135). The idea is that Foucault is actually advocating a "pluralist" view of truth that takes power relations to be more or less constitutive of truth, depending on the statement and context in question (compare Flynn 1988). If so, then Foucault may allow that some statements are less dependent for their truth on the current systems of power than others. Should this less extreme reading be correct, Foucault is less of a skeptic about truth than is frequently alleged.

#### Further Reading for Part IV

##### On Heidegger

Allen, B. 1993. Chapter 5 of *Truth in Philosophy*. Cambridge: Harvard University Press.

Gadamer, H. 1989. *Truth and Method*. 2nd ed. Trans. J. Weinsheimer and D. Marshall. New York: Crossroads Press.

Guignon, C. 1983. *Heidegger and the Problem of Knowledge*. Indianapolis: Hackett.

Heidegger, M. 1993. *Basic Writings*. Revised edition. Edited by D. Krell. San Francisco: HarperCollins.

Heidegger, M. 1982. *The Basic Problems of Phenomenology*. Trans. A. Hofstadter. Bloomington: Indiana University Press.

Heidegger, M. 1998. "Pathmarks." Trans. W. McNeill. Cambridge: Cambridge University Press. Contains his "Essence of Truth" and "The Essence of Reasons," as well as other important texts.

Heidegger, M. 1962. *Being and Time*. Trans. J. Macquarrie and E. Robinson. New York: Harper and Row.

Krell, D. 1986. Chapter 4 of *Intimations of Mortality: Time, Truth, and Finitude in Heidegger's Thinking of Being*. University Park: Pennsylvania State University Press.

Macomber, W. B. 1967. "Anatomy of Disillusion: Martin Heidegger's Notion of Truth." Evanston: Northwestern University Press.

Tugendhat, Ernst. 1969. "Heidegger's Idea of Truth." Reprinted in *Critical Heidegger*, edited by C. Macann. London: Routledge, 1996.

### On Foucault

Alcoff, L. 1995. Chapters 4 and 5 of *Real Knowing: New Versions of the Coherence Theory*. Ithaca: Cornell University Press.

Allen, B. 1993. Chapter 8 of *Truth in Philosophy*. Cambridge: Harvard University Press.

Flynn, T. 1988. "Foucault as Parrhesiast: His Last Course at the College de France." In *The Final Foucault*, edited by J. Bernauer and D. Rasmussen. Cambridge: MIT Press.

Foucault, M. 1980. *Power/Knowledge: Selected Interviews and Other Writings, 1972–1977*. Edited by C. Gordon. New York: Pantheon Books.

Foucault, M. 1982. *The Archaeology of Knowledge and The Discourse on Language*. Trans. A. M. Sheridan Smith. New York: Pantheon Books.

Foucault, M. 1988. "The Concern for Truth." In *Michel Foucault: Politics, Philosophy, Culture*, edited by Lawrence D. Kritzmann. New York: Routledge.

Habermas, J. 1989. *The Philosophical Discourse of Modernity*. Trans. F. Lawrence. Cambridge: MIT Press.

Rorty, R. 1986. "Foucault and Epistemology." In *Foucault: A Critical Reader*, edited by D. C. Hoy. Oxford: Basil Blackwell.

Taylor, C. 1986. "Foucault on Freedom and Truth." In *Foucault: A Critical Reader*, edited by D. C. Hoy. Oxford: Basil Blackwell.





---

## On the Essence of Truth

Martin Heidegger

Our topic is the *essence* of truth. The question regarding the essence of truth is not concerned with whether truth is a truth of practical experience or of economic calculation, the truth of a technical consideration or of political sagacity, or, in particular, a truth of scientific research or of artistic composition, or even the truth of thoughtful reflection or of cultic belief. The question of essence disregards all this and attends to the one thing that in general distinguishes every “truth” as truth.

Yet with this question concerning essence do we not soar too high into the void of generality which deprives all thinking of breath? Does not the extravagance of such questioning bring to light the groundlessness of all philosophy? A radical thinking that turns to what is actual must surely from the first insist bluntly on establishing the actual truth which today gives us a measure and a stand against the confusion of opinions and reckonings. In the face of this actual need what use is the question concerning the essence of truth, this “abstract” question that disregards everything actual? Is not the question of essence the most inessential and superfluous that could be asked?

No one can evade the evident certainty of these considerations. None can lightly neglect their compelling seriousness. But what is it that speaks in these considerations? “Sound” common sense. It harps on the demand for palpable utility and inveighs against knowledge of the essence of beings, which essential knowledge has long been called “philosophy.”<sup>1</sup>

Common sense has its own necessity; it asserts its rights with the weapon peculiarly suitable to it, namely, appeal to the “obviousness” of its claims and considerations. However, philosophy can never refute common sense, for the latter is deaf to the language of philosophy. Nor

may it even wish to do so, since common sense is blind to what philosophy sets before its essential vision.

Moreover, we ourselves remain within the sensibleness of common sense to the extent that we suppose ourselves to be secure in those multifiform “truths” of practical experience and action, of research, composition, and belief. We ourselves intensify that resistance which the “obvious” has to every demand made by what is questionable.

Therefore even if some questioning concerning truth is necessary, what we then demand is an answer to the question as to where we stand today. We want to know what our situation is today. We call for the goal which should be posited for man in and for his history. We want the actual “truth.” Well then—truth!

But in calling for the actual “truth” we must already know what truth as such means. Or do we know this only by “feeling” and “in a general way”? But is not such vague “knowing” and our indifference regarding it more desolate than sheer ignorance of the essence of truth?

## 1 The Usual Concept of Truth

What do we ordinarily understand by “truth”? This elevated yet at the same time worn and almost dulled word “truth” means what makes a true thing true. What is a true thing? We say, for example, “It is a true joy to cooperate in the accomplishment of this task.” We mean that it is purely and actually a joy. The true is the actual. Accordingly, we speak of true gold in distinction from false. False gold is not actually what it appears to be. It is merely a “semblance” and thus is not actual. What is not actual is taken to be the opposite of the actual. But what merely seems to be gold is nevertheless something actual. Accordingly, we say more precisely: actual gold is genuine gold. Yet both are “actual,” the circulating counterfeit no less than the genuine gold. What is true about genuine gold thus cannot be demonstrated merely by its actuality. The question recurs: what do “genuine” and “true” mean here? Genuine gold is that actual gold the actuality of which is in accordance [*in der Übereinstimmung steht*] with what, always and in advance, we “properly” mean by “gold.” Conversely, wherever we suspect false gold, we say: “Here something is not in accord” [*stimmt nicht*]. On the other

hand, we say of whatever is “as it should be”: “It is in accord.” The *matter* is in accord [*Die Sache stimmt*].

However, we call true not only an actual joy, genuine gold, and all beings of such kind, but also and above all we all true or false our statements about beings, which can themselves be genuine or not with regard to their kind, which can be thus or otherwise in their actuality. A statement is true if what it means and says is in accordance with the matter about which the statement is made. Here too we say, “It is in accord.” Now, though, it is not the matter that is in accord but rather the *proposition*.

The true, whether it be a matter or a proposition, is what accords, the accordant [*das Stimmende*]. Being true and truth here signify accord, and that in a double sense: on the one hand, the consonance [*Einstimmigkeit*] of a matter with what is supposed in advance regarding it and, on the other hand, the accordance of what is meant in the statement with the matter.

This dual character of the accord is brought to light by the traditional definition of truth: *veritas est adaequatio rei et intellectūs*. This can be taken to mean: truth is the correspondence [*Angleichung*] of the matter to knowledge. But it can also be taken as saying: truth is the correspondence of knowledge to the matter. Admittedly, the above definition is usually stated only in the formula *veritas est adaequatio intellectūs ad rem* [truth is the adequation of intellect to thing]. Yet truth so conceived, propositional truth, is possible only on the basis of material truth [*Sachwahrheit*], of *adaequatio rei ad intellectum* [adequation of thing to intellect]. Both concepts of the essence of *veritas* have continually in view a conforming to . . . [*Sichrichten nach . . .*], and hence think truth as *correctness* [*Richtigkeit*].

Nonetheless, the one is not the mere inversion of the other. On the contrary, in each case *intellectus* and *res* are thought differently. In order to recognize this we must trace the usual formula for the ordinary concept of truth back to its most recent (i.e., the medieval) origin. *Veritas* as *adaequatio rei ad intellectum* does not imply the later transcendental conception of Kant—possible only on the basis of the subjectivity of man’s essence—that “objects conform to our knowledge.” Rather, it implies the Christian theological belief that, with respect to what it is and whether it

is, a matter, as created (*ens creatum*), is only insofar as it corresponds to the idea preconceived in the *intellectus divinus*, i.e., in the mind of God, and thus measures up to the idea (is correct) and in this sense is “true.” The *intellectus humanus* too is an *ens creatum*. As a capacity bestowed upon man by God, it must satisfy its *idea*. But the understanding measures up to the idea only by accomplishing in its propositions the correspondence of what is thought to the matter, which in its turn must be in conformity with the *idea*. If all beings are “created,” the possibility of the truth of human knowledge is grounded in the fact that matter and proposition measure up to the idea in the same way and therefore are fitted to each other on the basis of the unity of the divine plan of creation. *Veritas* as *adaequatio rei (creandae) ad intellectum (divinum)* guarantees *veritas* as *adaequatio intellectūs (humani) ad rem (creatam)*. Throughout, *veritas* essentially implies *convenientia*, the coming of beings themselves, as created, into agreement with the Creator, an “accord” with regard to the way they are determined in the order of creation.

But this order, detached from the notion of creation, can also be represented in a general and indefinite way as a world-order. The theologically conceived order of creation is replaced by the capacity of all objects to be planned by means of a worldly reason [*Weltvernunft*] which supplies the law for itself and thus also claims that its procedure is immediately intelligible (what is considered “logical”). That the essence of propositional truth consists in the correctness of statements needs no further special proof. Even where an effort is made—with a conspicuous lack of success—to explain how correctness is to occur, it is already presupposed as being the essence of truth. Likewise, material truth always signifies the consonance of something at hand with the “rational” concept of its essence. The impression arises that this definition of the essence of truth is independent of the interpretation of the essence of the Being of all beings, which always includes a corresponding interpretation of the essence of man as the bearer and executor of *intellectus*. Thus the formula for the essence of truth (*veritas est adaequatio intellectūs et rei*) comes to have its general validity as something immediately evident to everyone. Under the domination of the obviousness which this concept of truth seems to have but which is hardly attended to as regards its essential grounds, it is considered equally obvious that truth has an opposite,

and that there is untruth. The untruth of the proposition (incorrectness) is the non-accordance of the statement with the matter. The untruth of the matter (non-genuineness) signifies non-agreement of a being with its essence. In each case untruth is conceived as a non-accord. The latter falls outside the essence of truth. Therefore when it is a question of comprehending the pure essence of truth, untruth, as such an opposite of truth, can be put aside.

But then is there any further need at all for a special unveiling of the essence of truth? Is not the pure essence of truth already adequately represented in the generally accepted concept, which is upset by no theory and is secured by its obviousness? Moreover, if we take the tracing back of propositional truth to material truth to be what in the first instance it shows itself to be, namely a theological explanation, and if we then keep the philosophical definition completely pure of all admixture of theology and limit the concept of truth to propositional truth, then we encounter an old—though not the oldest—tradition of thinking, according to which truth is the accordance (*homoïōsis*) of a statement (*logos*) with a matter (*pragma*). What is it about statements that here remains still worthy of question—granted that we know what is meant by accordance of a statement with the matter? Do we know that?

## 2 The Inner Possibility of Accordance

We speak of accordance in various senses. We say, for example, considering two five-mark coins lying on the table: they are in accordance with one another. They come into accord in the oneness of their outward appearance. Hence they have the latter in common, and thus they are in this regard alike. Furthermore, we speak of accordance whenever, for example, we state regarding one of the five-mark coins: this coin is round. Here the statement is in accordance with the thing. Now the relation obtains, not between thing and thing, but rather between a statement and a thing. But wherein are the thing and the statement supposed to be in accordance, considering that the relata are manifestly different in their outward appearance? The coin is made of metal. The statement is not material at all. The coin is round. The statement has nothing at all spatial about it. With the coin something can be purchased. The statement about

it is never a means of payment. But in spite of all their dissimilarity the above statement, as true, is in accordance with the coin. And according to the usual concept of truth this accord is supposed to be a correspondence. How can what is completely dissimilar, the statement, correspond to the coin? It would have to become the coin and in this way relinquish itself entirely. The statement never succeeds in doing that. The moment it did, it would no longer be able as a statement to be in accordance with the thing. In the correspondence the statement must remain—indeed even first become—what it is. In what does its essence, so thoroughly different from every thing, consist? How is the statement able to correspond to something else, the thing, precisely by persisting in its own essence?

Correspondence here cannot signify a thing-like approximation between dissimilar kinds of things. The essence of the correspondence is determined rather by the kind of relation that obtains between the statement and the thing. As long as this “relation” remains undetermined and is not grounded in its essence, all dispute over the possibility and impossibility, over the nature and degree, of the correspondence loses its way in a void. But the statement regarding the coin relates “itself” to this thing in that it presents [*vor-stellt*] it and says of the presented how, according to the particular perspective that guides it, it is disposed. What is stated by the presentative statement is said of the presented thing in just such manner *as* that thing, as presented, is. The “such-as” has to do with the presenting and its presented. Disregarding all “psychological” preconceptions as well as those of any “theory of consciousness,” to present here means to let the thing stand opposed as object. As thus placed, what stands opposed must traverse an open field of opposedness [*Entgegen*] and nevertheless must maintain its stand as a thing and show itself as something withstanding [*ein Ständiges*]. This appearing of the thing in traversing a field of opposedness takes place within an open region, the openness of which is not first created by the presenting but rather is only entered into and taken over as a domain of relatedness. The relation of the presentative statement to the thing is the accomplishment of that bearing [*Verhältnis*] which originally and always comes to prevail as a comportment [*Verhalten*]. But all comportment is distinguished by the fact that, standing in the open region, it adheres to something opened up

*as such*.<sup>2</sup> What is thus opened up, solely in this strict sense, was experienced early in Western thinking as “what is present” and for a long time has been named “being.”

Comportment stands open to beings. Every open relatedness is a comportment. Man’s open stance varies depending on the kind of beings and the way of comportment. All working and achieving, all action and calculation, keep within an open region within which beings, with regard to what they are and how they are, can properly take their stand and become capable of being said. This can occur only if beings present themselves along with the presentative statement so that the latter subordinates itself to the directive that it speak of beings *such-as* they are. In following such a directive the statement conforms to beings. Speech that directs itself accordingly is correct (true). What is thus said is the correct (the true).

A statement is invested with its correctness by the openness of comportment; for only through the latter can what is opened up really become the standard for the presentative correspondence. Open comportment must let itself be assigned this standard. This means that it must take over a pregiven standard for all presenting. This belongs to the openness of comportment. But if the correctness (truth) of statements becomes possible only through this openness of comportment, then what first makes correctness possible must with more original right be taken as the essence of truth.

Thus the traditional assignment of truth exclusively to statements as the sole essential locus of truth falls away. Truth does not originally reside in the proposition. But at the same time the question arises of the ground of the inner possibility of the open comportment which pregives a standard, which possibility alone lends to propositional correctness the appearance of fulfilling the essence of truth at all.

### 3 The Ground of the Possibility of Correctness

Whence does the presentative statement receive the directive to conform to the object and to accord by way of correctness? Why is this accord involved in determining the essence of truth? How can something like the accomplishment of a pregiven directedness occur? And how can the ini-



tiation into an accord occur? Only if this pregiving has already entered freely into an open region for something opened up which prevails there and which binds every presenting. To free oneself for a binding directness is possible only by *being free* for what is opened up in an open region. Such being free points to the heretofore uncomprehended essence of freedom. The openness of comportment as the inner condition of the possibility of correctness is grounded in freedom. *The essence of truth is freedom.*

But does not this proposition regarding the essence of correctness substitute one obvious item for another? In order to be able to carry out any act, and therefore one of presentative stating and even of according or not according with a “truth,” the actor must of course be free. However, the proposition in question does not really mean that an unconstrained act belongs to the execution of the statement, to its pronouncement and reception; rather, the proposition says that freedom is the *essence* of truth itself. In this connection “essence” is understood as the ground of the inner possibility of what is initially and generally admitted as known. Nevertheless, in the concept of freedom we do not think truth, and certainly not at all its essence. The proposition that the essence of truth (correctness of statements) is freedom must consequently seem strange.

To place the essence of truth in freedom—doesn’t this mean to submit truth to human caprice? Can truth be any more radically undermined than by being surrendered to the arbitrariness of this “waving reed”? What forced itself upon sound judgment again and again in the previous discussion now all the more clearly comes to light: truth is here driven back to the subjectivity of the human subject. Even if an objectivity is also accessible to this subject, still such objectivity remains along with subjectivity something human and at man’s disposal.

Certainly deceit and dissimulation, lies and deception, illusion and semblance—in short, all kinds of untruth—are ascribed to man. But of course untruth is also the opposite of truth. For this reason, as the non-essence of truth, it is appropriately excluded from the sphere of the question concerning the pure essence of truth. This human origin of untruth indeed only serves to confirm by contrast the essence of truth “in itself” as holding sway “beyond” man. Metaphysics regards such truth as the

imperishable and eternal, which can never be founded on the transitoriness and fragility that belong to man's essence. How then can the essence of truth still have its subsistence and its ground in human freedom?

Resistance to the proposition that the essence of truth is freedom is based on preconceptions, the most obstinate of which is that freedom is a property of man. The essence of freedom neither needs nor allows any further questioning. Everyone knows what man is.

#### 4 The Essence of Freedom

However, indication of the essential connection between truth as correctness and freedom uproots those preconceptions—granted of course that we are prepared for a transformation of thinking. Consideration of the essential connection between truth and freedom leads us to pursue the question of the essence of man in a regard which assures us an experience of a concealed essential ground of man (of *Dasein*), and in such a manner that the experience transposes us in advance into the originally essential domain of truth. But here it becomes evident also that freedom is the ground of the inner possibility of correctness only because it receives its own essence from the more original essence of uniquely essential truth. Freedom was first determined as freedom for what is opened up in an open region. How is this essence of freedom to be thought? That which is opened up, that to which a presentative statement as correct corresponds, are beings opened up in an open comportment. Freedom for what is opened up in an open region lets beings be the beings they are. Freedom now reveals itself as letting beings be.

Ordinarily we speak of letting be whenever, for example, we forgo some enterprise that has been planned. "We let something be" means we do not touch it again, we have nothing more to do with it. To let something be has here the negative sense of letting it alone, of renouncing it, of indifference and even neglect.

However, the phrase required now—to let beings be—does not refer to neglect and indifference but rather the opposite. To let be is to engage oneself with beings. On the other hand, to be sure, this is not to be understood only as the mere management, preservation, tending, and planning of the beings in each case encountered or sought out. To let be

—that is, to let beings be as the beings which they are—means to engage oneself with the open region and its openness into which every being comes to stand, bringing that openness, as it were, along with itself. Western thinking in its beginning conceived this open region as *ta alēthea*, the unconcealed. If we translate *alētheia* as “unconcealment” rather than “truth,” this translation is not merely more literal; it contains the directive to rethink the ordinary concept of truth in the sense of the correctness of statements and to think it back to that still uncomprehended disclosedness and disclosure of beings. To engage oneself with the disclosedness of beings is not to lose oneself in them; rather, such engagement withdraws in the face of beings in order that they might reveal themselves with respect to what and how they are and in order that presentative correspondence might take its standard from them. As this letting-be it exposes itself to beings as such and transposes all comportment into the open region. Letting-be, i.e., freedom, is intrinsically exposing, ek-sistent.<sup>3</sup> Considered in regard to the essence of truth, the essence of freedom manifests itself as exposure to the disclosedness of beings.

Freedom is not merely what common sense is content to let pass under this name: the caprice, turning up occasionally in our choosing, of inclining in this or that direction. Freedom is not mere absence of constraint with respect to what we can or cannot do. Nor is it on the other hand mere readiness for what is required and necessary (and so somehow a being). Prior to all this (“negative” and “positive” freedom), freedom is engagement in the disclosure of beings as such. Disclosedness itself is conserved in ek-sistent engagement, through which the openness of the open region, i.e., the “there” [*Da*], is what it is.

In *Da-sein* the essential ground, long ungrounded, on the basis of which man is able to ek-sist, is preserved for him. Here “existence” does not mean *existentia* in the sense of occurring or being at hand. Nor on the other hand does it mean, in an “existentiell” fashion, man’s moral endeavor in behalf of his “self,” based on his psychophysical constitution. Ek-sistence, rooted in truth as freedom, is exposure to the disclosedness of beings as such. Still uncomprehended, indeed, not even in need of an essential grounding, the ek-sistence of historical man begins at that moment when the first thinker takes a questioning stand with regard

to the unconcealment of beings by asking: what are beings? In this question unconcealment is experienced for the first time. Being as a whole reveals itself as *physis*, “nature,” which here does not yet mean a particular sphere of beings but rather beings as such as a whole, specifically in the sense of emerging presence [*aufgehendes Anwesen*]. History begins only when beings themselves are expressly drawn up into their unconcealment and conserved in it, only when this conservation is conceived on the basis of questioning regarding beings as such. The primordial disclosure of being as a whole, the question concerning beings as such, and the beginning of Western history are the same; they occur together in a “time” which, itself unmeasurable, first opens up the open region for every measure.

But if ek-sistent Da-sein, which lets beings be, sets man free for his “freedom” by first offering to his choice something possible (a being) and by imposing on him something necessary (a being), human caprice does not then have freedom at its disposal. Man does not “possess” freedom as a property. At best, the converse holds: freedom, ek-sistent, disclosive Da-sein, possesses man—so originally that only *it* secures for humanity that distinctive relatedness to being as a whole as such which first founds all history. Only ek-sistent man is historical. “Nature” has no history.

Freedom, understood as letting beings be, is the fulfillment and consummation of the essence of truth in the sense of the disclosure of beings. “Truth” is not a feature of correct propositions which are asserted of an “object” by a human “subject” and then “are valid” somewhere, in what sphere we know not; rather, truth is disclosure of beings through which an openness essentially unfolds [*west*]. All human comportment and bearing are exposed in its open region. Therefore man *is* in the manner of ek-sistence.

Because every mode of human comportment is in its own way open and plies itself to that toward which it comports itself, the restraint of letting-be, i.e., freedom, must have granted it its endowment of that inner directive for correspondence of presentation to beings. That man ek-sists now means that for historical humanity the history of its essential possibilities is conserved in the disclosure of beings as a whole. The rare and the simple decisions of history arise from the way the original essence of truth essentially unfolds.

However, because truth is in essence freedom, historical man can, in letting beings be, also *not* let beings be the beings which they are and as they are. Then beings are covered up and distorted. Semblance comes to power. In it the non-essence of truth comes to the fore. However, because ek-sistent freedom as the essence of truth is not a property of man; because on the contrary man ek-sists and so becomes capable of history only as the property of this freedom; the non-essence of truth cannot first arise subsequently from mere human incapacity and negligence. Rather, untruth must derive from the essence of truth. Only because truth and untruth are, *in essence*, *not* irrelevant to one another but rather belong together is it possible for a true proposition to enter into pointed opposition to the corresponding untrue proposition. The question concerning the essence of truth thus first reaches the original domain of what is at issue when, on the basis of a prior glimpse of the full essence of truth, it has included a consideration of untruth in its unveiling of that essence. Discussion of the non-essence of truth is not the subsequent filling of a gap but rather the decisive step toward an adequate posing of the *question* concerning the essence of truth. Yet how are we to comprehend the non-essence in the essence of truth? If the essence of truth is not exhausted by the correctness of statements, then neither can untruth be equated with the incorrectness of judgments.

## 5 The Essence of Truth

The essence of truth reveals itself as freedom. The latter is ek-sistent, disclosive letting beings be. Every mode of open comportment flourishes in letting beings be and in each case is a comportment to this or that being. As engagement in the disclosure of being as a whole as such, freedom has already attuned all comportment to being as a whole. However, being attuned (attunement)<sup>4</sup> can never be understood as “experience” and “feeling,” because it is thereby simply deprived of its essence. For here it is interpreted on the basis of something (“life” and “soul”) that can maintain the semblance of the title of essence only as long as it bears in itself the distortion and misinterpretation of being attuned. Being attuned, i.e., ek-sistent exposedness to beings as a whole, can be “experienced” and “felt” only because the “man who experiences,” without

being aware of the essence of the attunement, is always engaged in being attuned in a way that discloses beings as a whole. Every mode of historical man's comportment—whether accentuated or not, whether understood or not—is attuned and by this attunement is drawn up into beings as a whole. The openedness of being as a whole does not coincide with the sum of all immediately familiar beings. On the contrary: where beings are not very familiar to man and are scarcely and only roughly known by science, the openedness of beings as a whole can prevail more essentially than it can where the familiar and well-known has become boundless, and nothing is any longer able to withstand the business of knowing, since technical mastery over things bears itself without limit. Precisely in the leveling and planing of this omniscience, this mere knowing, the openedness of beings gets flattened out into the apparent nothingness of what is no longer even a matter of indifference but rather is simply forgotten.

Letting beings be, which is an attuning, a bringing into accord, prevails throughout and anticipates all the open comportment that flourishes in it. Man's comportment is brought into definite accord throughout by the openedness of being as a whole. However, from the point of view of everyday calculations and preoccupations this "as a whole" appears to be incalculable and incomprehensible. It cannot be understood on the basis of the beings opened up in any given case, whether they belong to nature or to history. Although it ceaselessly brings everything into definite accord, still it remains indefinite, indeterminable; it then coincides for the most part with what is most fleeting and most unconsidered. However, what brings into accord is not nothing but rather a concealing of beings as a whole. Precisely because letting be always lets beings be in a particular comportment which relates to them and thus discloses them, it conceals beings as a whole. Letting-be is intrinsically at the same time a concealing. In the ek-sistent freedom of Da-sein a concealing of being as a whole comes to pass [*ereignet sich*]. Here there *is* concealment.

## 6 Untruth as Concealing

Concealment deprives *alētheia* of disclosure yet does not render it *sterēsis* (privation); rather, concealment preserves what is most proper to *alētheia*

as its own. Considered with respect to truth as disclosedness, concealment is then undisclosedness and accordingly the untruth that is most proper to the essence of truth. The concealment of beings as a whole does not first show up subsequently as a consequence of the fact that knowledge of beings is always fragmentary. The concealment of beings as a whole, untruth proper, is older than every openedness of this or that being. It is also older than letting-be itself which in disclosing already holds concealed and comports itself toward concealing. What conserves letting-be in this relatedness to concealing? Nothing less than the concealing of what is concealed as a whole, of beings as such, i.e., the mystery; not a particular mystery regarding this or that, but rather the one mystery—that, in general, mystery (the concealing of what is concealed) as such holds sway throughout man's Da-sein.

In letting beings as a whole be, which discloses and at the same time conceals, it happens that concealing appears as what is first of all concealed. Insofar as it ek-sists, Da-sein conserves the first and broadest undisclosedness, untruth proper. The proper non-essence of truth is the mystery. Here non-essence does not yet have the sense of inferiority to essence in the sense of what is general (*koimon, genos*), its *possibilitas* and the ground of its possibility. Non-essence is here what in such a sense would be a pre-essential essence. But “non-essence” means at first and for the most part the deformation of that already inferior essence. Indeed, in each of these significations the non-essence remains always in its own way essential to the essence and never becomes inessential in the sense of irrelevant. But to speak of non-essence and untruth in this manner goes very much against the grain of ordinary opinion and looks like a dragging up of forcibly contrived *paradoxa*. Because it is difficult to eliminate this impression, such a way of speaking, paradoxical only for ordinary *doxa* (opinion), is to be renounced. But surely for those who know about such matters the “non-” of the primordial non-essence of truth, as untruth, points to the still unexperienced domain of the truth of Being (not merely of beings).

As letting beings be, freedom is intrinsically the resolutely open bearing that does not close up in itself.<sup>5</sup> All comportment is grounded in this bearing and receives from it directedness toward beings and disclosure of

them. Nevertheless, this bearing toward concealing conceals itself in the process, letting a forgottenness of the mystery take precedence and disappearing in it. Certainly man takes his bearings [*verhält sich*] constantly in his comportment toward beings; but for the most part he acquiesces in this or that being and its particular openedness. Man clings to what is readily available and controllable even where ultimate matters are concerned. And if he sets out to extend, change, newly assimilate, or secure the openedness of the beings pertaining to the most various domains of his activity and interest, then he still takes his directives from the sphere of readily available intentions and needs.

However, to reside in what is readily available is intrinsically not to let the concealing of what is concealed hold sway. Certainly among readily familiar things there are also some that are puzzling, unexplained, undecided, questionable. But these self-certain questions are merely transitional, intermediate points in our movement within the readily familiar and thus not essential. Wherever the concealment of beings as a whole is conceded only as a limit that occasionally announces itself, concealing as a fundamental occurrence has sunk into forgottenness.

But the forgotten mystery of Dasein is not eliminated by the forgottenness; rather, the forgottenness bestows on the apparent disappearance of what is forgotten a peculiar presence [*Gegenwart*]. By disavowing itself in and for forgottenness, the mystery leaves historical man in the sphere of what is readily available to him, leaves him to his own resources. Thus left, humanity replenishes its "world" on the basis of the latest needs and aims, and fills out that world by means of proposing and planning. From these man then takes his standards, forgetting being as a whole. He persists in them and continually supplies himself with new standards, yet without considering either the ground for taking up standards or the essence of what gives the standard. In spite of his advance to new standards and goals, man goes wrong as regards the essential genuineness of his standards. He is all the more mistaken the more exclusively he takes himself, as subject, to be the standard for all beings. The inordinate forgetfulness of humanity persists in securing itself by means of what is readily available and always accessible. This persistence has its unwitting support in that *bearing* by which Dasein not only ek-sists but also at the



same time *in-sists*, i.e., holds fast to what is offered by beings, as if they were open of and in themselves.

As *ek-sistent*, *Dasein* is *insistent*. Even in insistent existence the mystery holds sway, but as the forgotten and hence “inessential” essence of truth.

## 7 Untruth as Errancy

As insistent, man is turned toward the most readily available beings. But he insists only by being already *ek-sistent*, since, after all, he takes beings as his standard. However, in taking its standard, humanity is turned away from the mystery. The insistent turning toward what is readily available and the *ek-sistent* turning away from the mystery belong together. They are one and the same. Yet turning toward and away from is based on a turning to and fro proper to *Dasein*. Man’s flight from the mystery toward what is readily available, onward from one current thing to the next, passing the mystery by—this is *erring*.<sup>6</sup>

Man errs. Man does not merely stray into errancy. He is always astray in errancy, because as *ek-sistent* he *in-sists* and so already is caught in errancy. The errancy through which man strays is not something which, as it were, extends alongside man like a ditch into which he occasionally stumbles; rather errancy belongs to the inner constitution of the *Da-sein* into which historical man is admitted. Errancy is the free space for that turning in which insistent *ek-sistence* adroitly forgets and mistakes itself constantly anew. The concealing of the concealed being as a whole holds sway in that disclosure of specific beings, which, as forgottenness of concealment, becomes errancy.

Errancy is the essential counter-essence to the primordial essence of truth. Errancy opens itself up as the open region for every opposite to essential truth. Errancy is the open site for and ground of *error*. Error is not just an isolated mistake but rather the realm (the domain) of the history of those entanglements in which all kinds of *erring* get interwoven.

In conformity with its openness and its relatedness to beings as a whole, every mode of comportment has its mode of *erring*. Error extends

from the most ordinary wasting of time, making a mistake, and miscalculating, to going astray and venturing too far in one's essential attitudes and decisions. However, what is ordinarily and even according to the teachings of philosophy recognized as error, incorrectness of judgments and falsity of knowledge, is only one mode of erring and, moreover, the most superficial one. The errancy in which any given segment of historical humanity must proceed for its course to be errant is essentially connected with the openness of Dasein. By leading him astray, errancy dominates man through and through. But, as leading astray, errancy at the same time contributes to a possibility that man is capable of drawing up from his ek-sistence—the possibility that, by experiencing errancy itself and by not mistaking the mystery of Da-sein, he *not* let himself be led astray.

Because man's in-sistent ek-sistence proceeds in errancy, and because errancy as leading astray always oppresses in some manner or other and is formidable on the basis of this oppression of the mystery, specifically as something forgotten, in the ek-sistence of his Dasein man is *especially* subjected to the rule of the mystery and the oppression of errancy. He is in the *needful condition of being constrained* by the one and the other. The full essence of truth, including its most proper non-essence, keeps Dasein in need by this perpetual turning to and fro. Dasein is a turning into need. From man's Dasein and from it alone arises the disclosure of necessity and, as a result, the possibility of being transposed into what is inevitable.

The disclosure of beings as such is simultaneously and intrinsically the concealing of being as a whole. In the simultaneity of disclosure and concealing errancy holds sway. Errancy and the concealing of what is concealed belong to the primordial essence of truth. Freedom, conceived on the basis of the in-sistent ek-sistence of Dasein, is the essence of truth (in the sense of the correctness of presenting) only because freedom itself originates from the primordial essence of truth, the rule of the mystery in errancy. Letting beings be takes its course in open comportment. However, letting beings as such be as a whole occurs in a way befitting its essence only when from time to time it gets taken up in its primordial essence. Then resolute openness toward the mystery [*Ent-schlossenheit*

*zum Geheimnis*] is under way into errancy as such. Then the question of the essence of truth gets asked more originally. Then the ground of the intertwining of the essence of truth with the truth of essence reveals itself. The glimpse into the mystery out of errancy is a question—in the sense of that unique question of what being as such is as a whole. This questioning thinks the question of the *Being* of beings, a question that is essentially misleading and thus in its manifold meaning is still not mastered. The thinking of Being, from which such questioning primordially originates, has since Plato been understood as “philosophy” and later received the title “metaphysics.”

## 8 Philosophy and the Question of Truth

In the thinking of Being the liberation of man for ek-sistence, the liberation that grounds history, is put into words. These are not just the “expression” of an opinion but are always already the ably conserved articulation of the truth of being as a whole. How many have ears for these words matters not. Who those are that can hear them determines man’s standpoint in history. However, in the same period in which the beginning of philosophy takes place, the *marked* domination of common sense (sophistry) also begins.

Sophistry appeals to the unquestionable character of the beings that are opened up and interprets all thoughtful questioning as an attack on, an unfortunate irritation of, common sense.

However, what philosophy is according to the estimation of common sense, which is quite justified in its own domain, does not touch on the essence of philosophy, which can be determined only on the basis of relatedness to the original truth of being as such as a whole. But because the full essence of truth contains the non-essence and above all holds sway as concealing, philosophy as a questioning into this truth is intrinsically discordant. Philosophical thinking is gentle releasement that does not renounce the concealment of being as a whole. Philosophical thinking is especially the stern and resolute openness that does not disrupt the concealing but entreats its unbroken essence into the open region of understanding and thus into its own truth.

In the gentle sternness and stern gentleness with which it lets being as such be as a whole, philosophy becomes a questioning which does not cling solely to beings yet which also can allow no externally imposed decree. Kant presaged this innermost need that thinking has. For he says of philosophy:

Here philosophy is seen in fact to be placed in a precarious position which is supposed to be stable—although neither in heaven nor on earth is there anything on which it depends or on which it is based. It is here that it has to prove its integrity as the keeper of its laws [*Selbhalterin ihrer Gesetze*], not as the mouthpiece of laws secretly communicated to it by some implanted sense or by who knows what tutelary nature. (*Grundlegung zur Metaphysik der Sitten. Werke*, Akademieausgabe IV, 425.)

With this essential interpretation of philosophy, Kant, whose work introduces the final turning of Western metaphysics, envisions a domain which to be sure he could understand only on the basis of his fundamental metaphysical position, founded on subjectivity, and which he had to understand as the keeping of its laws. This essential view of the determination of philosophy nevertheless goes far enough to renounce every subjugation of philosophical thinking, the most destitute kind of which lets philosophy still be of value as an “expression” of “culture” (Spengler) and as an ornament of productive mankind.

However, whether philosophy as “keeper of its laws” fulfills its primordially decisive essence, or whether it is not itself first of all kept and appointed to its task as keeper by the truth of that to which its laws pertain—this depends on the primordially with which the original essence of truth becomes essential for thoughtful questioning.

The present undertaking takes the question of the essence of truth beyond the confines of the ordinary definition provided in the usual concept of essence and helps us to consider whether the question of the essence of truth must not be, at the same time and even first of all, the question concerning the truth of essence. But in the concept of “essence” philosophy thinks Being. In tracing the inner possibility of the correctness of statements back to the ek-sistent freedom of letting-be as its “ground,” likewise in pointing to the essential commencement of this ground in concealing and in errancy, we want to show that the essence of truth is not the empty “generality” of an “abstract” universality but rather that

which, self-concealing, is unique in the unremitting history of the disclosure of the “meaning” of what we call Being—what we for a long time have been accustomed to considering only as being as a whole.

## 9 Note

The question of the essence of truth arises from the question of the truth of essence. In the former question essence is understood initially in the sense of whatness (*quidditas*) or material content (*realitas*), whereas truth is understood as a characteristic of knowledge. In the question of the truth of essence, essence is understood verbally; in this word, remaining still within metaphysical presentation, Being is thought as the difference that holds sway between Being and beings. Truth signifies sheltering that lightens [*lichtendes Bergen*] as the basic characteristic of Being. The question of the essence of truth finds its answer in the proposition *the essence of truth is the truth of essence*. After our explanation it can easily be seen that the proposition does not merely reverse the word order so as to conjure the specter of paradox. The subject of the proposition—if this unfortunate grammatical category may still be used at all—is the truth of essence. Sheltering that lightens is—i.e., lets essentially unfold—accordance between knowledge and beings. The proposition is not dialectical. It is no proposition at all in the sense of a statement. The answer to the question of the essence of truth is the saying of a turning [*die Sage einer Kehre*] within the history of Being. Because sheltering that lightens belongs to it, Being appears primordially in the light of concealing withdrawal. The name of this lighting [*Lichtung*] is *alētheia*.

Already in the original project the lecture “On the Essence of Truth” was to have been completed by a second lecture “On the Truth of Essence.” The latter failed for reasons that are now indicated in the “Letter on Humanism.”

The decisive question (in *Being and Time*, 1927) of the meaning, i.e., of the project-domain, i.e., of the openness, i.e., of the truth of Being and not merely of beings, remains intentionally undeveloped. Our thinking apparently remains on the path of metaphysics. Nevertheless, in its decisive steps, which lead from truth as correctness to ek-sistent freedom, and

from the latter to truth as concealing and as errancy, it accomplishes a change in the questioning that belongs to the overcoming of metaphysics. The thinking attempted in the lecture comes to fulfillment in the essential experience that a nearness to the truth of Being is first prepared for historical man on the basis of the Da-sein into which man can enter. Every kind of anthropology and all subjectivity of man as subject is not merely left behind—as it was already in *Being and Time*—and the truth of Being sought as the ground of a transformed historical position; rather, the movement of the lecture is such that it sets out to think from this other ground (Da-sein). The course of the questioning is intrinsically the way of a thinking which, instead of furnishing representations and concepts, experiences and tries itself as a transformation of its relatedness to Being.

### Notes

This translation of *Vom Wesen der Wahrheit* is by John Sallis. The German text is contained in Martin Heidegger, *Wegmarken* (Frankfurt am Main: Vittorio Klostermann Verlag, 1967), pp. 73–97. This translation is based on the fourth edition of the essay, published by Klostermann in 1961.

1. Throughout the translation *das Seiende* is rendered as “being” or “beings,” *ein Seiendes* as “a being,” *Sein* as “Being,” *das Seiende im Ganzen* as either “being as a whole” or “beings as a whole” depending on the context.
2. The text reads, “ein Offenbares *als ein solches*.” In ordinary German *offenbar* means “evident,” “manifest.” However, the context which it has here through its link with “open region” (*das Offene*), “open stance” (*Offenständigkeit*), and “openness” (*Offenheit*) already suggests the richer sense that the word has for Heidegger: that of something’s being so opened up as to reveal itself, to be manifest (as, for example, a flower in bloom), in contrast to something’s being so closed or sealed up within itself that it conceals itself.
3. This variant of the word *Existenz* indicates the ecstatic character of freedom, its standing outside itself.
4. The text reads, “*Die Gestimmtheit (Stimmung) . . .*” *Stimmung* refers not only to the kind of attunement which a musical instrument receives by being tuned but also to the kind of attunement that constitutes a mood or a disposition of Dasein. The important etymological connection between *Stimmung* and the various formations based on *stimmen* (to accord) is not retained in the translation.
5. “Resolutely open bearing” seeks to translate *das entschlossene Verhältnis*. *Entschlossen* is usually rendered as “resolute,” but such a translation fails to retain the word’s structural relation to *verschlossen*, “closed” or “shut up.” Significantly, this connection is what makes it possible for Heidegger to transform the sense of the word: he takes the prefix as a privation rather than as indicating establish-

ment of the condition designated by the word to which it is affixed. Thus, as the text here makes quite clear, *entschlossen* signifies just the opposite of that kind of “resolve” in which one makes up his mind in such fashion as to close off all other possibilities: it is rather a kind of keeping *un-closed*.—Tr.

6. “To err” may translate *irren* only if it is understood in its root sense derived from the Latin *errare*, “to wander from the right way,” and only secondarily in the sense “to fall into error.”

---

## Truth and Power

Michel Foucault

The important thing here, I believe, is that truth isn't outside power, or lacking in power: contrary to a myth whose history and functions would repay further study, truth isn't the reward of free spirits, the child of protracted solitude, nor the privilege of those who have succeeded in liberating themselves. Truth is a thing of this world: it is produced only by virtue of multiple forms of constraint. And it induces regular effects of power. Each society has its régime of truth, its 'general politics' of truth: that is, the types of discourse which it accepts and makes function as true; the mechanisms and instances which enable one to distinguish true and false statements, the means by which each is sanctioned; the techniques and procedures accorded value in the acquisition of truth; the status of those who are charged with saying what counts as true.

In societies like ours, the 'political economy' of truth is characterised by five important traits. 'Truth' is centred on the form of scientific discourse and the institutions which produce it; it is subject to constant economic and political incitement (the demand for truth, as much for economic production as for political power); it is the object, under diverse forms, of immense diffusion and consumption (circulating through apparatuses of education and information whose extent is relatively broad in the social body, notwithstanding certain strict limitations); it is produced and transmitted under the control, dominant if not exclusive, of a few great political and economic apparatuses (university, army, writing, media); lastly, it is the issue of a whole political debate and social confrontation ('ideological' struggles).

It seems to me that what must now be taken into account in the intellectual is not the 'bearer of universal values'. Rather, it's the person



occupying a specific position—but whose specificity is linked, in a society like ours, to the general functioning of an apparatus of truth. In other words, the intellectual has a three-fold specificity: that of his class position (whether as petty-bourgeois in the service of capitalism or ‘organic’ intellectual of the proletariat); that of his conditions of life and work, linked to his condition as an intellectual (his field of research, his place in a laboratory, the political and economic demands to which he submits or against which he rebels, in the university, the hospital, ect.); lastly, the specificity of the politics of truth in our societies. And it’s with this last factor that his position can take on a general significance and that his local, specific struggle can have effects and implications which are not simply professional or sectoral. The intellectual can operate and struggle at the general level of that régime of truth which is so essential to the structure and functioning of our society. There is a battle ‘for truth’, or at least ‘around truth’—it being understood once again that by truth I do not mean ‘the ensemble of truths which are to be discovered and accepted’, but rather ‘the ensemble of rules according to which the true and the false are separated and specific effects of power attached to the true’, it being understood also that it’s not a matter of a battle ‘on behalf’ of the truth, but of a battle about the status of truth and the economic and political role it plays. It is necessary to think of the political problems of intellectuals not in terms of ‘science’ and ‘ideology’, but in terms of ‘truth’ and ‘power’. And thus the question of the professionalisation of intellectuals and the division between intellectual and manual labour can be envisaged in a new way.

All this must seem very confused and uncertain. Uncertain indeed, and what I am saying here is above all to be taken as a hypothesis. In order for it to be a little less confused, however, I would like to put forward a few ‘propositions’—not firm assertions, but simply suggestions to be further tested and evaluated.

‘Truth’ is to be understood as a system of ordered procedures for the production, regulation, distribution, circulation and operation of statements.

‘Truth’ is linked in a circular relation with systems of power which produce and sustain it, and to effects of power which it induces and which extend it. A ‘régime’ of truth.

This régime is not merely ideological or superstructural; it was a condition of the formation and development of capitalism. And it's this same régime which, subject to certain modifications, operates in the socialist countries (I leave open here the question of China, about which I know little).

The essential political problem for the intellectual is not to criticise the ideological contents supposedly linked to science, or to ensure that his own scientific practice is accompanied by a correct ideology, but that of ascertaining the possibility of constituting a new politics of truth. The problem is not changing people's consciousnesses—or what's in their heads—but the political, economic, institutional régime of the production of truth.

It's not a matter of emancipating truth from every system of power (which would be a chimera, for truth is already power) but of detaching the power of truth from the forms of hegemony, social, economic and cultural, within which it operates at the present time.

The political question, to sum up, is not error, illusion, alienated consciousness or ideology; it is truth itself. Hence the importance of Nietzsche.



# V

---

## Tarski's Theory and Its Importance



---

# Tarski's Theory and Its Importance:

## Introduction

Most philosophers would agree that Alfred Tarski's work on truth has been immensely influential. But that is where agreement generally ends, for the point of Tarski's theory, its applicability to philosophical concerns, and its very nature are all deeply disputed.

When discussing Tarski's importance, it is crucial to distinguish two influential elements of his work that are often conflated. The first is what Tarski called his *material-adequacy condition* for a theory of truth. This is Tarski's famous "Convention T" or "schema T." Tarski's idea was that it was a minimal condition of any theory of truth that it entail all sentences of the following form:

(T)  $X$  is true if and only if  $p$ .

Here " $p$ " is a variable for a sentence and " $X$ " is a name of for that sentence, and the "if and only if" denotes an extensional equivalence. Instances of (T) therefore include the following:

"Roses are red" is true if and only if roses are red.

"Violets are blue" is true if and only if violets are blue.

Tarski was not the first to note that biconditionals of this form capture something very basic about our concept of truth (see, e.g., Ramsey, chap. 18). But he was the first to employ schema (T) as a test for truth theories. Since schema (T) captured the most basic fact about our concept of truth, Tarski reasoned, any adequate theory of truth must logically entail every instance of this schema in the language where the predicate is being defined.

It is important to realize, however, that (T) is not Tarski's *definition* of truth.<sup>1</sup> His definition is the second and more important element of Tarski's work on truth that has received attention. Nonetheless, while (T)

does not comprise Tarski's own theory of truth, he did believe that each instance of (T) is "a partial definition of truth"; each instance defines truth for the sentence in question, e.g., "Roses are red." Thus, if the language contained only the sentences used in the examples above, we could give a complete definition of truth by simply conjoining the respective instances of (T). But Tarski was interested in languages where it was possible to have an infinite number of sentences, and therefore he aimed to supply one general definition of truth that would be extensionally equivalent to the logical conjunction of an infinite number of instances of (T).

Tarski had three goals in presenting a theory of truth. First, he obviously wanted a theory that would meet the material-adequacy condition. Second, Tarski wanted to make the concept of truth physically respectable. In the 1930s, when Tarski first developed his theory, verificationist philosophers were openly suspicious of semantic notions like truth. Like the verificationists, Tarski subscribed to physicalism, or the view that any concept worth having can be defined completely in terms that refer only to physical and mathematical entities. As a result, Tarski set himself the task to define truth in this way for certain formal languages, or languages thought to be adequate for the physical sciences and mathematics.

Finally, he wanted a theory that was immune from destruction by the Liar Paradox. We can informally state the paradox by considering (1):

(1) This sentence is false.

From this seemingly innocuous statement a contradiction follows. For (1) must be either true or false, but if it is either, it is both true and false. After all, if (1) is true, then it is false, because it says that it is false. But if it is false, then it must be true, since once again, it says that it is false. So (1) is both true and false, which is a contradiction.<sup>2</sup>

Tarski argues that natural languages, like English, are subject to the paradox because they are "semantically closed"; that is, they contain semantic predicates like "true" or "false" that can apply to the language's own sentences. Tarski therefore restricts his definition to semantically open languages, or artificial languages whose semantic predicates apply only to the sentences of languages other than itself. Liar sentences

like (1) can't be expressed in such a language. But this means that if we are going to define truth for this semantically open language, we must use another semantically open language to do so. That is, we must distinguish between the language for which we want to define truth (the "object language") and the language in which we express that definition (the "metalinguage"). When we state our definition in the metalanguage, we must use words like "true" to mean "true-in-L" (or "true-in-the-object-language"). In short, to avoid the liar paradox, Tarski believed that a definition of "true" must not be expressed in the language for which the concept is being defined. We must always climb up to a metalanguage to define truth for the language below.

With this qualification in place, Tarski defines truth as a relation, which he calls "satisfaction," that links expressions and objects: "A sentence is true when satisfied by all objects and false otherwise."<sup>3</sup> His definition relies on the idea that while a language has a potentially infinite number of sentences, those sentences are constructed from a finite vocabulary. The rough idea is first to define the basic elements of the language and then to build up the definition from there by means of a procedure known as recursion. A recursive definition consists of one or more clauses that specify the most basic members of a particular set, followed by further clauses that show how other members of the set are built out of the more basic members. As long as there is a finite number of types of basic members of the set and a finite number of ways these can be combined to form new (nonbasic) members, an infinite number of nonbasic members of the set can be defined by this procedure. In his definition, Tarski took the basic expressions of the languages he was interested in to be quantifiers and sentential functions (or open sentences such as "*x* loves *y*" or "*y* is a philosopher"). We can then say that satisfaction relates sentential functions to objects, roughly as follows. If "*x* is red" is the sentential function and a rose is the object, then the rose *satisfies* "*x* is red" just when the rose is red. Simplifying greatly, what Tarski showed is that once we define the simplest types of sentential functions in this way, we can use the truth-functional operators ("and," "or," "not," etc.) and existential quantifiers ("all" and "some") to define "satisfies" and hence "true" for any complete sentence (such as "There is a red rose" or "All roses are red") in the language.



Two features of Tarski's definition in particular need highlighting. First, it must be stressed that the languages that concerned Tarski were entirely formal languages, such as the languages of deductive logic. Not only are such languages semantically open in the sense explained above; they contain none of the ambiguity and vagueness that so riddle our ordinary language. Secondly, it is crucial to Tarski's definition that it is not a general definition of "true" in any language *L*, but a definition of "true-in-*L*<sub>1</sub>," "true-in-*L*<sub>2</sub>," and so on. This is a consequence not only of Tarski's meta/object language distinction but also of the fact that his definitions rely on extensional definitions of the basic expressions (the predicates, essentially) of the language. Languages with different basic expressions or predicates will therefore entail a distinct definition of "true."

These two features of Tarski's theory have caused many philosophers to wonder whether it is at all relevant to the philosophical problem of defining the ordinary concept of truth. After all, the languages *we* speak are messy, rife with ambiguity, and subject to paradox. Further, a definition of "true-in-*L*<sub>1</sub>," "true-in-*L*<sub>2</sub>," etc., may seem to miss the point of defining "true," in the same way that definitions of "the proper legal ruling-on-Wednesday" and the "the proper legal ruling-on-Thursday" don't give us a definition of "proper legal ruling" (see Blackburn 1984, 266–267). We want to know what all the relativized definitions have in common; that is, we want a general definition of "true."

In his seminal paper on Tarski, Hartry Field (chap. 16) argues that Tarski's definition can be thought of as consisting of two parts. The first part defines what it is for a name to denote an object and for a predicate to apply to some class of objects. The second part then recursively defines "true" in terms of object denotation and predicate application. According to Field, Tarski's definition, when thought of in this way, can be applied to natural languages, so long as we take the truth bearers to be sentence tokens, or individual utterances. But if we do so, Field argues, we must admit that Tarski did not succeed in incorporating truth into a physicalist picture of the world. Instead, he only reduced truth to other unexplicated semantic notions. Field's point is that the definitions of satisfaction for Tarski's basic expressions do not reduce or explain the

crucial notions of satisfaction or predicate application but simply list the satisfaction (or application) conditions for each atomic open sentence (or predicate) in the language. He argues that to reduce truth to purely physical terms, we must go on to give more than a “listlike” definition of satisfaction, denotation, or predicate application. A promising possibility, Field suggests, is to employ causal theories of reference to this end.

Scott Soames, in his influential paper (which is in large part a reply to Field) argues that Tarski’s theory is consistent with physicalism and that it is indeed ontologically neutral with regard to substantive metaphysical debates. Soames notes that a central part of Field’s argument against Tarski rests on the assumption that linguistic expressions (such as “true”) have their semantic properties only in virtue of how those expressions are used by the speakers of the language. Field sees languages as comprised of utterances whose semantic features are speaker-dependent. In reply, Soames argues that we should instead see Tarski’s definitions as applying to languages seen as abstract objects that have their semantic properties essentially. If we see languages in this way, Tarski’s definition appears to be a purely mathematical or formal one. (Compare Field’s response included in chap. 21).

Field’s and Soames’ papers also encourage different answers to one of the most fundamental questions in the Tarskian literature, namely, whether Tarski’s theory should be taken as a correspondence theory of truth. Tarski himself argues that his theory, in so far as it meets the material adequacy condition, does accord with the fundamental intuitions behind the correspondence theory. But it is unclear whether this entails that the theory itself can be called a correspondence account of truth. If we follow Field, then we might think that Tarski’s theory amounts to (or should amount to) a correspondence theory in which “correspondence” is cashed out in terms of causal relations between our uses of words and the world (see the introduction to part I). On the other hand, Soames’ interpretation takes it that what is right about Tarski’s approach is its “deflationary character.” On this view, which is also Quine’s (chap. 20), the important lesson of Tarski’s theory is that truth is a useful but essentially unmysterious notion. It is not the key to understanding how our thought relates to the world.

## Notes

1. Indeed, he wouldn't have understood how an open sentence like (T) could be a definition.
2. The liar paradox is frustrating because while (1) is odd-sounding, it is more difficult than it looks to say what exactly is wrong with it. It seems like a perfectly meaningful sentence, and although it is self-referring, that in itself can't be a reason to reject it as meaningless—"This sentence is in English" is also a self-referring sentence. Further, the premises that we used to derive a contradiction from (1) seem eminently respectable: a sentence is true when and only when things are as it says they are; and every sentence is either true or false. Given these facts, it is not surprising that the liar paradox remains a formidable problem for any theory of truth. We don't even need a self-referring sentence to generate the paradox, as consideration of the following pair of sentences illustrates: The next sentence is false. The last sentence is true. Moreover, Kripke (1975) has shown that the paradox can arise in the context of "normal" conversation.
3. As Tarski notes in his essay in this volume, various technical details require that the definition actually be put in terms of sequences of objects, e.g., truth is satisfaction by all sequences of objects (see Tarski 1933a). For a user-friendly explanation of the point of this provision, and of Tarski's theory in general, see Kirkham, 1992, chap. 5.

## Further Reading for Part V

- Black, M. 1948. "The Semantic Conception of Truth." *Analysis* 8: 49–63.
- Blackburn, S. 1984. *Spreading the Word*. Oxford: Oxford University Press.
- Davidson, D. 1984. *Inquiries into Truth and Meaning*. New York: Oxford University Press.
- Etchemendy, J. 1988. "Tarski on Truth and Logical Consequence." *Journal of Symbolic Logic* 53: 51–79.
- Etchemendy, J. 1990. *The Concept of Logical Consequence*. Cambridge: Harvard University Press.
- Fox, J. 1989. "What Were Tarski's Truth Definitions For?" *History and Philosophy of Logic* 10: 165–179.
- Haack, S. 1978. *Philosophy of Logics*. Cambridge: Cambridge University Press.
- Hintikka, J. 1975. "A Counterexample to Tarski-Type Truth Conditions As Applied to Natural Languages." *Philosophia* 5: 207–212.
- Keuth, H. "Tarski's Definition of Truth and the Correspondence Theory." *Philosophy of Science* 45: 420–430.
- Kirkham, R. 1992. Chaps. 5, 6, and 9 of *Theories of Truth*. Cambridge: MIT Press.
- Kripke, S. 1975. "Outline of a Theory of Truth." *Journal of Philosophy* 72: 690–716.

- McDowell, J. 1978. "Physicalism and Primitive Denotation: Field on Tarski." *Erkenntnis* 13: 137–152.
- Popper, K. 1974. "Some Philosophical Comments on Tarski's Theory of Truth." In *Proceedings of the Tarski Symposium*, pp. 400–402. Providence: American Mathematical Society.
- Quine, W. V. O. 1970. *Philosophy of Logic*. Englewood Cliffs: Prentice-Hall.
- Soames, S. 1999. *Understanding Truth*. New York: Oxford University Press.
- Tarski, A. 1933a. "The Concept of Truth in Formalized Languages." In his *Logic, Semantics, Metamathematics*, 2nd ed., trans. J. H. Woodger. Indianapolis: Hackett, 1983.
- Tarski, A. 1933b. "The Establishment of Scientific Semantics." In his *Logic, Semantics, and Metamathematics*, 2nd ed., trans. J. H. Woodger. Indianapolis: Hackett, 1983.
- Tarski, A. 1969. "Truth and Proof." *Scientific American* 220: 63–67.



---

## The Semantic Conception of Truth and the Foundations of Semantics

Alfred Tarski

This paper consists of two parts; the first has an expository character, and the second is rather polemical.

In the first part I want to summarize in an informal way the main results of my investigations concerning the definition of truth and the more general problem of the foundations of semantics. These results have been embodied in a work which appeared in print several years ago.<sup>1</sup> Although my investigations concern concepts dealt with in classical philosophy, they happen to be comparatively little known in philosophical circles, perhaps because of their strictly technical character. For this reason I hope I shall be excused for taking up the matter once again.<sup>2</sup>

Since my work was published, various objections, of unequal value, have been raised to my investigations; some of these appeared in print, and others were made in public and private discussions in which I took part.<sup>3</sup> In the second part of the paper I should like to express my views regarding these objections. I hope that the remarks which will be made in this context will not be considered as purely polemical in character, but will be found to contain some constructive contributions to the subject.

In the second part of the paper I have made extensive use of material graciously put at my disposal by Dr. Marja Kokoszyńska (University of Lwów). I am especially indebted and grateful to Professors Ernest Nagel (Columbia University) and David Rynin (University of California, Berkeley) for their help in preparing the final text and for various critical remarks.

## I Exposition

### 1 The Main Problem—A Satisfactory Definition of Truth

Our discussion will be centered around the notion<sup>4</sup> of *truth*. The main problem is that of giving a *satisfactory definition* of this notion, i.e., a definition which is *materially adequate* and *formally correct*. But such a formulation of the problem, because of its generality, cannot be considered unequivocal, and requires some further comments.

In order to avoid any ambiguity, we must first specify the conditions under which the definition of truth will be considered adequate from the material point of view. The desired definition does not aim to specify the meaning of a familiar word used to denote a novel notion; on the contrary, it aims to catch hold of the actual meaning of an old notion. We must then characterize this notion precisely enough to enable anyone to determine whether the definition actually fulfills its task.

Secondly, we must determine on what the formal correctness of the definition depends. Thus, we must specify the words or concepts which we wish to use in defining the notion of truth; and we must also give the formal rules to which the definition should conform. Speaking more generally, we must describe the formal structure of the language in which the definition will be given.

The discussion of these points will occupy a considerable portion of the first part of the paper.

### 2 The Extension of the Term “True”

We begin with some remarks regarding the extension of the concept of truth which we have in mind here.

The predicate “*true*” is sometimes used to refer to psychological phenomena such as judgments or beliefs, sometimes to certain physical objects, namely, linguistic expressions and specifically sentences, and sometimes to certain ideal entities called “propositions.” By “sentence” we understand here what is usually meant in grammar by “declarative sentence”; as regards the term “proposition,” its meaning is notoriously a subject of lengthy disputations by various philosophers and logicians, and it seems never to have been made quite clear and unambiguous. For

several reasons it appears most convenient to *apply the term "true" to sentences*, and we shall follow this course.<sup>5</sup>

Consequently, we must always relate the notion of truth, like that of a sentence, to a specific language; for it is obvious that the same expression which is a true sentence in one language can be false or meaningless in another.

Of course, the fact that we are interested here primarily in the notion of truth for sentences does not exclude the possibility of a subsequent extension of this notion to other kinds of objects.

### 3 The Meaning of the Term "True"

Much more serious difficulties are connected with the problem of the meaning (or the intension) of the concept of truth.

The word "*true*," like other words from our everyday language, is certainly not unambiguous. And it does not seem to me that the philosophers who have discussed this concept have helped to diminish its ambiguity. In works and discussions of philosophers we meet many different conceptions of truth and falsity, and we must indicate which conception will be the basis of our discussion.

We should like our definition to do justice to the intuitions which adhere to the *classical Aristotelian conception of truth*—intuitions which find their expression in the well-known words of Aristotle's *Metaphysics*:

To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, or of what is not that it is not, is true.

If we wished to adapt ourselves to modern philosophical terminology, we could perhaps express this conception by means of the familiar formula:

The truth of a sentence consists in its agreement with (or correspondence to) reality.

(For a theory of truth which is to be based upon the latter formulation the term "correspondence theory" has been suggested.)

If, on the other hand, we should decide to extend the popular usage of the term "*designate*" by applying it not only to names, but also to sentences, and if we agreed to speak of the designata of sentences as "states



of affairs,” we could possibly use for the same purpose the following phrase:

A sentence is true if it designates an existing state of affairs.<sup>6</sup>

However, all these formulations can lead to various misunderstandings, for none of them is sufficiently precise and clear (though this applies much less to the original Aristotelian formulation than to either of the others); at any rate, none of them can be considered a satisfactory definition of truth. It is up to us to look for a more precise expression of our intuitions.

#### 4 A Criterion for the Material Adequacy of the Definition<sup>7</sup>

Let us start with a concrete example. Consider the sentence “*snow is white*.” We ask the question under what conditions this sentence is true or false. It seems clear that if we base ourselves on the classical conception of truth, we shall say that the sentence is true if snow is white, and that it is false if snow is not white. Thus, if the definition of truth is to conform to our conception, it must imply the following equivalence:

The sentence “snow is white” is true if, and only if, snow is white.

Let me point out that the phrase “*snow is white*” occurs on the left side of this equivalence in quotation marks, and on the right without quotation marks. On the right side we have the sentence itself, and on the left the name of the sentence. Employing the medieval logical terminology we could also say that on the right side the words “*snow is white*” occur in *suppositio formalis*, and on the left in *suppositio materialis*. It is hardly necessary to explain why we must have the name of the sentence, and not the sentence itself, on the left side of the equivalence. For, in the first place, from the point of view of the grammar of our language, an expression of the form “*X is true*” will not become a meaningful sentence if we replace in it ‘*X*’ by a sentence or by anything other than a name—since the subject of a sentence may be only a noun or an expression functioning like a noun. And, in the second place, the fundamental conventions regarding the use of any language require that in any utterance we make about an object it is the name of the object which must be employed, and not the object itself. In consequence, if we wish to say

something about a sentence, for example, that it is true, we must use the name of this sentence, and not the sentence itself.<sup>8</sup>

It may be added that enclosing a sentence in quotation marks is by no means the only way of forming its name. For instance, by assuming the usual order of letters in our alphabet, we can use the following expression as the name (the description) of the sentence "*snow is white*":

the sentence constituted by three words, the first of which consists of the 19th, 14th, 15th, and 23rd letters, the second of the 9th and 19th letters, and the third of the 23rd, 8th, 9th, 20th, and 5th letters of the English alphabet

We shall now generalize the procedure which we have applied above. Let us consider an arbitrary sentence; we shall replace it by the letter '*p*.' We form the name of this sentence and we replace it by another letter, say '*X*.' We ask now what is the logical relation between the two sentences "*X is true*" and '*p*.' It is clear that from the point of view of our basic conception of truth these sentences are equivalent. In other words, the following equivalence holds:

(T) *X is true if, and only if, p.*

We shall call any such equivalence (with '*p*' replaced by any sentence of the language to which the word "*true*" refers, and '*X*' replaced by a name of this sentence) an "*equivalence of the form (T).*"

Now at last we are able to put into a precise form the conditions under which we will consider the usage and the definition of the term "*true*" as adequate from the material point of view: we wish to use the term "*true*" in such a way that all equivalences of the form (T) can be asserted, and *we shall call a definition of truth "adequate" if all these equivalences follow from it.*

It should be emphasized that neither the expression (T) itself (which is not a sentence, but only a schema of a sentence) nor any particular instance of the form (T) can be regarded as a definition of truth. We can only say that every equivalence of the form (T) obtained by replacing '*p*' by a particular sentence, and '*X*' by a name of this sentence, may be considered a partial definition of truth, which explains wherein the truth of this one individual sentence consists. The general definition has to be, in a certain sense, a logical conjunction of all these partial definitions.

(The last remark calls for some comments. A language may admit the construction of infinitely many sentences; and thus the number of partial definitions of truth referring to sentences of such a language will also be infinite. Hence to give our remark a precise sense we should have to explain what is meant by a “logical conjunction of infinitely many sentences”; but this would lead us too far into technical problems of modern logic.)

## 5 Truth as a Semantic Concept

I should like to propose the name “*the semantic conception of truth*” for the conception of truth which has just been discussed.

*Semantics* is a discipline which, speaking loosely, *deals with certain relations between expressions of a language and the objects* (or “states of affairs”) “*referred to*” by those expressions. As typical examples of semantic concepts we may mention the concepts of *designation*, *satisfaction*, and *definition* as these occur in the following examples:

the expression “the father of his country” designates (denotes) George Washington

snow satisfies the sentential function (the condition) “ $x$  is white”

the equation “ $2 \cdot x = 1$ ” defines (uniquely determines) the number  $1/2$

While the words “*designates*,” “*satisfies*,” and “*defines*” express relations (between certain expressions and the objects “referred to” by these expressions), the word “*true*” is of a different logical nature: it expresses a property (or denotes a class) of certain expressions, viz., of sentences. However, it is easily seen that all the formulations which were given earlier and which aimed to explain the meaning of this word (cf. Sections 3 and 4) referred not only to sentences themselves, but also to objects “talked about” by these sentences, or possibly to “states of affairs” described by them. And, moreover, it turns out that the simplest and the most natural way of obtaining an exact definition of truth is one which involves the use of other semantic notions, e.g., the notion of satisfaction. It is for these reasons that we count the concept of truth which is discussed here among the concepts of semantics, and the problem of defining truth proves to be closely related to the more general problem of setting up the foundations of theoretical semantics.

It is perhaps worth while saying that semantics as it is conceived in this paper (and in former papers of the author) is a sober and modest discipline which has no pretensions of being a universal patent-medicine for all the ills and diseases of mankind, whether imaginary or real. You will not find in semantics any remedy for decayed teeth or illusions of grandeur or class conflicts. Nor is semantics a device for establishing that everyone except the speaker and his friends is speaking nonsense.

From antiquity to the present day the concepts of semantics have played an important role in the discussions of philosophers, logicians, and philologists. Nevertheless, these concepts have been treated for a long time with a certain amount of suspicion. From a historical standpoint, this suspicion is to be regarded as completely justified. For although the meaning of semantic concepts as they are used in everyday language seems to be rather clear and understandable, still all attempts to characterize this meaning in a general and exact way miscarried. And what is worse, various arguments in which these concepts were involved, and which seemed otherwise quite correct and based upon apparently obvious premises, led frequently to paradoxes and antinomies. It is sufficient to mention here the *antinomy of the liar*, Richard's *antinomy of definability* (by means of a finite number of words), and Grelling-Nelson's *antinomy of heterological terms*.<sup>9</sup>

I believe that the method which is outlined in this paper helps to overcome these difficulties and assures the possibility of a consistent use of semantic concepts.

## 6 Languages with a Specified Structure

Because of the possible occurrence of antinomies, the problem of specifying the formal structure and the vocabulary of a language in which definitions of semantic concepts are to be given becomes especially acute; and we turn now to this problem.

There are certain general conditions under which the structure of a language is regarded as *exactly specified*. Thus, to specify the structure of a language, we must characterize unambiguously the class of those words and expressions which are to be considered *meaningful*. In particular, we must indicate all words which we decide to use without defining them, and which are called "*undefined (or primitive) terms*"; and we must give

the so-called *rules of definition* for introducing new or *defined terms*. Furthermore, we must set up criteria for distinguishing within the class of expressions those which we call "*sentences*." Finally, we must formulate the conditions under which a sentence of the language can be *asserted*. In particular, we must indicate all *axioms* (or *primitive sentences*), i.e., those sentences which we decide to assert without proof; and we must give the so-called *rules of inference* (or *rules of proof*) by means of which we can deduce new asserted sentences from other sentences which have been previously asserted. Axioms, as well as sentences deduced from them by means of rules of inference, are referred to as "*theorems*" or "*provable sentences*."

If in specifying the structure of a language we refer exclusively to the form of the expressions involved, the language is said to be *formalized*. In such a language theorems are the only sentences which can be asserted.

At the present time the only languages with a specified structure are the formalized languages of various systems of deductive logic, possibly enriched by the introduction of certain non-logical terms. However, the field of application of these languages is rather comprehensive; we are able, theoretically, to develop in them various branches of science, for instance, mathematics and theoretical physics.

(On the other hand, we can imagine the construction of languages which have an exactly specified structure without being formalized. In such a language the assertability of sentences, for instance, may depend not always on their form, but sometimes on other, non-linguistic factors. It would be interesting and important actually to construct a language of this type, and specifically one which would prove to be sufficient for the development of a comprehensive branch of empirical science; for this would justify the hope that languages with specified structure could finally replace everyday language in scientific discourse.)

*The problem of the definition of truth obtains a precise meaning and can be solved in a rigorous way only for those languages whose structure has been exactly specified.* For other languages—thus, for all natural, "spoken" languages—the meaning of the problem is more or less vague, and its solution can have only an approximate character. Roughly speaking, the approximation consists in replacing a natural language (or a portion of it in which we are interested) by one whose structure is

exactly specified, and which diverges from the given language “as little as possible.”

## 7 The Antinomy of the Liar

In order to discover some of the more specific conditions which must be satisfied by languages in which (or for which) the definition of truth is to be given, it will be advisable to begin with a discussion of that antinomy which directly involves the notion of truth, namely, the antinomy of the liar.

To obtain this antinomy in a perspicuous form,<sup>10</sup> consider the following sentence:

The sentence printed in this paper on p. 339, l. 11, is not true.

For brevity we shall replace the sentence just stated by the letter ‘s.’

According to our convention concerning the adequate usage of the term “*true*,” we assert the following equivalence of the form (T):

- (1) ‘s’ is true if, and only if, the sentence printed in this paper on p. 339, l. 11, is not true.

On the other hand, keeping in mind the meaning of the symbol ‘s,’ we establish empirically the following fact:

- (2) ‘s’ is identical with the sentence printed in this paper on p. 339, l. 11.

Now, by a familiar law from the theory of identity (Leibniz’s law), it follows from (2) that we may replace in (1) the expression “*the sentence printed in this paper on p. 339, l. 11*” by the symbol “‘s.’” We thus obtain what follows:

- (3) ‘s’ is true if, and only if, ‘s’ is not true.

In this way we have arrived at an obvious contradiction.

In my judgment, it would be quite wrong and dangerous from the standpoint of scientific progress to depreciate the importance of this and other antinomies, and to treat them as jokes or sophistries. It is a fact that we are here in the presence of an absurdity, that we have been compelled to assert a false sentence (since (3), as an equivalence between two contradictory sentences, is necessarily false). If we take our work seriously, we cannot be reconciled with this fact. We must discover its cause, that is

to say, we must analyze premises upon which the antinomy is based; we must then reject at least one of these premises, and we must investigate the consequences which this has for the whole domain of our research.

It should be emphasized that antinomies have played a preeminent role in establishing the foundations of modern deductive sciences. And just as class-theoretical antinomies, and in particular Russell's antinomy (of the class of all classes that are not members of themselves), were the starting point for the successful attempts at a consistent formalization of logic and mathematics, so the antinomy of the liar and other semantic antinomies give rise to the construction of theoretical semantics.

## 8 The Inconsistency of Semantically Closed Languages<sup>7</sup>

If we now analyze the assumptions which lead to the antinomy of the liar, we notice the following:

I. We have implicitly assumed that the language in which the antinomy is constructed contains, in addition to its expressions, also the names of these expressions, as well as semantic terms such as the term "*true*" referring to sentences of this language; we have also assumed that all sentences which determine the adequate usage of this term can be asserted in the language. A language with these properties will be called "*semantically closed*."

II. We have assumed that in this language the ordinary laws of logic hold.

III. We have assumed that we can formulate and assert in our language an empirical premise such as the statement (2) which has occurred in our argument.

It turns out that the assumption (III) is not essential, for it is possible to reconstruct the antinomy of the liar without its help.<sup>11</sup> But the assumptions (I) and (II) prove essential. Since every language which satisfies both of these assumptions is inconsistent, we must reject at least one of them.

It would be superfluous to stress here the consequences of rejecting the assumption (II), that is, of changing our logic (supposing this were possible) even in its more elementary and fundamental parts. We thus consider only the possibility of rejecting the assumption (I). Accordingly, we decide *not to use any language which is semantically closed* in the sense given.

This restriction would of course be unacceptable for those who, for reasons which are not clear to me, believe that there is only one “genuine” language (or, at least, that all “genuine” languages are mutually translatable). However, this restriction does not affect the needs or interests of science in any essential way. The languages (either the formalized languages or—what is more frequently the case—the portions of everyday language) which are used in scientific discourse do not have to be semantically closed. This is obvious in case linguistic phenomena and, in particular, semantic notions do not enter in any way into the subject matter of a science; for in such a case the language of this science does not have to be provided with any semantic terms at all. However, we shall see in the next section how semantically closed languages can be dispensed with even in those scientific discussions in which semantic notions are essentially involved.

The problem arises as to the position of everyday language with regard to this point. At first blush it would seem that this language satisfies both assumptions (I) and (II), and that therefore it must be inconsistent. But actually the case is not so simple. Our everyday language is certainly not one with an exactly specified structure. We do not know precisely which expressions are sentences, and we know even to a smaller degree which sentences are to be taken as assertible. Thus the problem of consistency has no exact meaning with respect to this language. We may at best only risk the guess that a language whose structure has been exactly specified and which resembles our everyday language as closely as possible would be inconsistent.

## **9 Object-Language and Meta-language**

Since we have agreed not to employ semantically closed languages, we have to use two different languages in discussing the problem of the definition of truth and, more generally, any problems in the field of semantics. The first of these languages is the language which is “talked about” and which is the subject matter of the whole discussion; the definition of truth which we are seeking applies to the sentences of this language. The second is the language in which we “talk about” the first language, and in terms of which we wish, in particular, to construct the definition of



truth for the first language. We shall refer to the first language as “*the object-language*,” and to the second as “*the meta-language*.”

It should be noticed that these terms “object-language” and “meta-language” have only a relative sense. If, for instance, we become interested in the notion of truth applying to sentences, not of our original object-language, but of its meta-language, the latter becomes automatically the object-language of our discussion; and in order to define truth for this language, we have to go to a new meta-language—so to speak, to a meta-language of a higher level. In this way we arrive at a whole hierarchy of languages.

The vocabulary of the meta-language is to a large extent determined by previously stated conditions under which a definition of truth will be considered materially adequate. This definition, as we recall, has to imply all equivalences of the form (T):

(T)  $X$  is true if, and only if,  $p$ .

The definition itself and all the equivalences implied by it are to be formulated in the meta-language. On the other hand, the symbol ‘ $p$ ’ in (T) stands for an arbitrary sentence of our object-language. Hence it follows that every sentence which occurs in the object-language must also occur in the meta-language; in other words, the meta-language must contain the object-language as a part. This is at any rate necessary for the proof of the adequacy of the definition—even though the definition itself can sometimes be formulated in a less comprehensive meta-language which does not satisfy this requirement.

(The requirement in question can be somewhat modified, for it suffices to assume that the object-language can be translated into the meta-language; this necessitates a certain change in the interpretation of the symbol ‘ $p$ ’ in (T). In all that follows we shall ignore the possibility of this modification.)

Furthermore, the symbol ‘ $X$ ’ in (T) represents the name of the sentence which ‘ $p$ ’ stands for. We see therefore that the meta-language must be rich enough to provide possibilities of constructing a name for every sentence of the object-language.

In addition, the meta-language must obviously contain terms of a general logical character, such as the expression “if, and only if.”<sup>12</sup>

It is desirable for the meta-language not to contain any undefined terms except such as are involved explicitly or implicitly in the remarks above, i.e.: terms of the object-language; terms referring to the form of the expressions of the object-language, and used in building names for these expressions; and terms of logic. In particular, we desire *semantic terms* (referring to the object-language) *to be introduced into the meta-language only by definition*. For, if this postulate is satisfied, the definition of truth, or of any other semantic concept, will fulfill what we intuitively expect from every definition; that is, it will explain the meaning of the term being defined in terms whose meaning appears to be completely clear and unequivocal. And, moreover, we have then a kind of guarantee that the use of semantic concepts will not involve us in any contradictions.

We have no further requirements as to the formal structure of the object-language and the meta-language; we assume that it is similar to that of other formalized languages known at the present time. In particular, we assume that the usual formal rules of definition are observed in the meta-language.

## 10 Conditions for a Positive Solution of the Main Problem

Now, we have already a clear idea both of the conditions of material adequacy to which the definition of truth is subjected, and of the formal structure of the language in which this definition is to be constructed. Under these circumstances the problem of the definition of truth acquires the character of a definite problem of a purely deductive nature.

The solution of the problem, however, is by no means obvious, and I would not attempt to give it in detail without using the whole machinery of contemporary logic. Here I shall confine myself to a rough outline of the solution and to the discussion of certain points of a more general interest which are involved in it.

The solution turns out to be sometimes positive, sometimes negative. This depends upon some formal relations between the object-language and its meta-language; or, more specifically, upon the fact whether the meta-language in its logical part is “*essentially richer*” than the object-language or not. It is not easy to give a general and precise definition of this notion of “essential richness.” If we restrict ourselves to languages

based on the logical theory of types, the condition for the meta-language to be “essentially richer” than the object-language is that it contain variables of a higher logical type than those of the object-language.

If the condition of “essential richness” is not satisfied, it can usually be shown that an interpretation of the meta-language in the object-language is possible; that is to say, with any given term of the meta-language a well-determined term of the object-language can be correlated in such a way that the assertible sentences of the one language turn out to be correlated with assertible sentences of the other. As a result of this interpretation, the hypothesis that a satisfactory definition of truth has been formulated in the meta-language turns out to imply the possibility of reconstructing in that language the antinomy of the liar; and this in turn forces us to reject the hypothesis in question.

(The fact that the meta-language, in its non-logical part, is ordinarily more comprehensive than the object-language does not affect the possibility of interpreting the former in the latter. For example, the names of expressions of the object-language occur in the meta-language, though for the most part they do not occur in the object-language itself; but, nevertheless, it may be possible to interpret these names in terms of the object-language.)

Thus we see that the condition of “essential richness” is necessary for the possibility of a satisfactory definition of truth in the meta-language. If we want to develop the theory of truth in a meta-language which does not satisfy this condition, we must give up the idea of defining truth with the exclusive help of those terms which were indicated above (in Section 8). We have then to include the term “*true*,” or some other semantic term, in the list of undefined terms of the meta-language, and to express fundamental properties of the notion of truth in a series of axioms. There is nothing essentially wrong in such an axiomatic procedure, and it may prove useful for various purposes.<sup>13</sup>

It turns out, however, that this procedure can be avoided. For *the condition of the “essential richness” of the meta-language proves to be, not only necessary, but also sufficient for the construction of a satisfactory definition of truth*; i.e., if the meta-language satisfies this condition, the notion of truth can be defined in it. We shall now indicate in general terms how this construction can be carried through.

## 11 The Construction (in Outline) of the Definition<sup>14</sup>

A definition of truth can be obtained in a very simple way from that of another semantic notion, namely, of the notion of *satisfaction*.

Satisfaction is a relation between arbitrary objects and certain expressions called "*sentential functions*." These are expressions like "*x is white*," "*x is greater than y*," etc. Their formal structure is analogous to that of sentences; however, they may contain the so-called free variables (like '*x*' and '*y*' in "*x is greater than y*"), which cannot occur in sentences.

In defining the notion of a sentential function in formalized languages, we usually apply what is called a "recursive procedure"; i.e., we first describe sentential functions of the simplest structure (which ordinarily presents no difficulty), and then we indicate the operations by means of which compound functions can be constructed from simpler ones. Such an operation may consist, for instance, in forming the logical disjunction or conjunction of two given functions, i.e., by combining them by the word "*or*" or "*and*." A sentence can now be defined simply as a sentential function which contains no free variables.

As regards the notion of satisfaction, we might try to define it by saying that given objects satisfy a given function if the latter becomes a true sentence when we replace in it free variables by names of given objects. In this sense, for example, snow satisfies the sentential function "*x is white*" since the sentence "*snow is white*" is true. However, apart from other difficulties, this method is not available to us, for we want to use the notion of satisfaction in defining truth.

To obtain a definition of satisfaction we have rather to apply again a recursive procedure. We indicate which objects satisfy the simplest sentential functions; and then we state the conditions under which given objects satisfy a compound function—assuming that we know which objects satisfy the simpler functions from which the compound one has been constructed. Thus, for instance, we say that given numbers satisfy the logical disjunction "*x is greater than y or x is equal to y*" if they satisfy at least one of the functions "*x is greater than y*" or "*x is equal to y*."

Once the general definition of satisfaction is obtained, we notice that it applies automatically also to those special sentential functions which contain no free variables, i.e., to sentences. It turns out that for a sentence only two cases are possible: a sentence is either satisfied by all objects, or

by no objects. Hence we arrive at a definition of truth and falsehood simply by saying that *a sentence is true if it is satisfied by all objects, and false otherwise*.<sup>15</sup>

(It may seem strange that we have chosen a roundabout way of defining the truth of a sentence, instead of trying to apply, for instance, a direct recursive procedure. The reason is that compound sentences are constructed from simpler sentential functions, but not always from simpler sentences; hence no general recursive method is known which applies specifically to sentences.)

From this rough outline it is not clear where and how the assumption of the “essential richness” of the meta-language is involved in the discussion; this becomes clear only when the construction is carried through in a detailed and formal way.<sup>16</sup>

## 12 Consequences of the Definition

The definition of truth which was outlined above has many interesting consequences.

In the first place, the definition proves to be not only formally correct, but also materially adequate (in the sense established in Section 4); in other words, it implies all equivalences of the form (T). In this connection it is important to notice that the conditions for the material adequacy of the definition determine uniquely the extension of the term “*true*.” Therefore, every definition of truth which is materially adequate would necessarily be equivalent to that actually constructed. The semantic conception of truth gives us, so to speak, no possibility of choice between various non-equivalent definitions of this notion.

Moreover, we can deduce from our definition various laws of a general nature. In particular, we can prove with its help the *laws of contradiction and of excluded middle*, which are so characteristic of the Aristotelian conception of truth; i.e., we can show that one and only one of any two contradictory sentences is true. These semantic laws should not be identified with the related logical laws of contradiction and excluded middle; the latter belong to the sentential calculus, i.e., to the most elementary part of logic, and do not involve the term “*true*” at all.

Further important results can be obtained by applying the theory of truth to formalized languages of a certain very comprehensive class of

mathematical disciplines; only disciplines of an elementary character and a very elementary logical structure are excluded from this class. It turns out that for a discipline of this class *the notion of truth never coincides with that of provability*; for all provable sentences are true, but there are true sentences which are not provable.<sup>17</sup> Hence it follows further that every such discipline is consistent, but incomplete; that is to say, of any two contradictory sentences at most one is provable, and—what is more—there exists a pair of contradictory sentences neither of which is provable.<sup>18</sup>

### 13 Extension of the Results to Other Semantic Notions

Most of the results at which we arrived in the preceding sections in discussing the notion of truth can be extended with appropriate changes to other semantic notions, for instance, to the notion of satisfaction (involved in our previous discussion), and to those of *designation* and *definition*.

Each of these notions can be analyzed along the lines followed in the analysis of truth. Thus, criteria for an adequate usage of these notions can be established; it can be shown that each of these notions, when used in a semantically closed language according to those criteria, leads necessarily to a contradiction;<sup>19</sup> a distinction between the object-language and the meta-language becomes again indispensable; and the “essential richness” of the meta-language proves in each case to be a necessary and sufficient condition for a satisfactory definition of the notion involved. Hence the results obtained in discussing one particular semantic notion apply to the general problem of the foundations of theoretical semantics.

Within theoretical semantics we can define and study some further notions, whose intuitive content is more involved and whose semantic origin is less obvious; we have in mind, for instance, the important notions of *consequence*, *synonymy*, and *meaning*.<sup>20</sup>

We have concerned ourselves here with the theory of semantic notions related to an individual object-language (although no specific properties of this language have been involved in our arguments). However, we could also consider the problem of developing *general semantics* which applies to a comprehensive class of object-languages. A considerable part of our previous remarks can be extended to this general problem; how-

ever, certain previous remarks can be extended to this general problem; however, certain new difficulties arise in this connection, which will not be discussed here. I shall merely observe that the axiomatic method (mentioned in Section 10) may prove the most appropriate for the treatment of the problem.<sup>21</sup>

## II Polemical Remarks

### 14 Is the Semantic Conception of Truth the “Right” One?

I should like to begin the polemical part of the paper with some general remarks.

I hope nothing which is said here will be interpreted as a claim that the semantic conception of truth is the “right” or indeed the “only possible” one. I do not have the slightest intention to contribute in any way to those endless, often violent discussions on the subject: “What is the right conception of truth?”<sup>22</sup> I must confess I do not understand what is at stake in such disputes; for the problem itself is so vague that no definite solution is possible. In fact, it seems to me that the sense in which the phrase “the right conception” is used has never been made clear. In most cases one gets the impression that the phrase is used in an almost mystical sense based upon the belief that every word has only one “real” meaning (a kind of Platonic or Aristotelian idea), and that all the competing conceptions really attempt to catch hold of this one meaning; since, however, they contradict each other, only one attempt can be successful, and hence only one conception is the “right” one.

Disputes of this type are by no means restricted to the notion of truth. They occur in all domains where—instead of an exact, scientific terminology—common language with its vagueness and ambiguity is used; and they are always meaningless, and therefore in vain.

It seems to me obvious that the only rational approach to such problems would be the following: We should reconcile ourselves with the fact that we are confronted, not with one concept, but with several different concepts which are denoted by one word; we should try to make these concepts as clear as possible (by means of definition, or of an axiomatic procedure, or in some other way); to avoid further confusions, we should agree to use different terms for different concepts; and then we may pro-

ceed to a quiet and systematic study of all concepts involved, which will exhibit their main properties and mutual relations.

Referring specifically to the notion of truth, it is undoubtedly the case that in philosophical discussions—and perhaps also in everyday usage—some incipient conceptions of this notion can be found that differ essentially from the classical one (of which the semantic conception is but a modernized form). In fact, various conceptions of this sort have been discussed in the literature, for instance, the pragmatic conception, the coherence theory, etc.<sup>6</sup>

It seems to me that none of these conceptions have been put so far in an intelligible and unequivocal form. This may change, however; a time may come when we find ourselves confronted with several incompatible, but equally clear and precise, conceptions of truth. It will then become necessary to abandon the ambiguous usage of the word “*true*,” and to introduce several terms instead, each to denote a different notion. Personally, I should not feel hurt if a future world congress of the “theoreticians of truth” should decide—by a majority of votes—to reserve the word “*true*” for one of the non-classical conceptions, and should suggest another word, say, “*frue*,” for the conception considered here. But I cannot imagine that anybody could present cogent arguments to the effect that the semantic conception is “wrong” and should be entirely abandoned.

## 15 Formal Correctness of the Suggested Definition of Truth

The specific objections which have been raised to my investigations can be divided into several groups; each of these will be discussed separately.

I think that practically all these objections apply, not to the special definition I have given, but to the semantic conception of truth in general. Even those which were leveled against the definition actually constructed could be related to any other definition which conforms to this conception.

This holds, in particular, for those objections which concern the formal correctness of the definition. I have heard a few objections of this kind; however, I doubt very much whether anyone of them can be treated seriously.

As a typical example let me quote in substance such an objection.<sup>23</sup> In formulating the definition we use necessarily sentential connectives, i.e.,



expressions like “*if ... , then*,” “*or*,” etc. They occur in the definiens; and one of them, namely, the phrase “*if, and only if*” is usually employed to combine the definiendum with the definiens. However, it is well known that the meaning of sentential connectives is explained in logic with the help of the words “*true*” and “*false*”; for instance, we say that an equivalence, i.e., a sentence of the form “*p if, and only if, q*,” is true if either both of its members, i.e., the sentences represented by ‘*p*’ and ‘*q*,’ are true or both are false. Hence the definition of truth involves a vicious circle.

If this objection were valid, no formally correct definition of truth would be possible; for we are unable to formulate any compound sentence without using sentential connectives, or other logical terms defined with their help. Fortunately, the situation is not so bad.

It is undoubtedly the case that a strictly deductive development of logic is often preceded by certain statements explaining the conditions under which sentences of the form “*if p, then q*,” etc., are considered true or false. (Such explanations are often given schematically, by means of the so-called truth-tables.) However, these statements are outside of the system of logic, and should not be regarded as definitions of the terms involved. They are not formulated in the language of the system, but constitute rather special consequences of the definition of truth given in the meta-language. Moreover, these statements do not influence the deductive development of logic in any way. For in such a development we do not discuss the question whether a given sentence is true, we are only interested in the problem whether it is provable.<sup>24</sup>

On the other hand, the moment we find ourselves within the deductive system of logic—or of any discipline based upon logic, e.g., of semantics—we either treat sentential connectives as undefined terms, or else we define them by means of other sentential connectives, but never by means of semantic terms like “*true*” or “*false*.” For instance, if we agree to regard the expressions “*not*” and “*if ... , then*” (and possibly also “*if, and only if*”) as undefined terms, we can define the term “*or*” by stating that a sentence of the form “*p or q*” is equivalent to the corresponding sentence of the form “*if not p, then q*.” The definition can be formulated, e.g., in the following way:

( $p$  or  $q$ ) if, and only if, (if not  $p$ , then  $q$ )

This definition obviously contains no semantic terms.

However, a vicious circle in definition arises only when the definiens contains either the term to be defined itself, or other terms defined with its help. Thus we clearly see that the use of sentential connectives in defining the semantic term "*true*" does not involve any circle.

I should like to mention a further objection which I have found in the literature and which seems also to concern the formal correctness, if not of the definition of truth itself, then at least of the arguments which lead to this definition.<sup>25</sup>

The author of this objection mistakenly regards scheme (T) (from Section 4) as a definition of truth. He charges this alleged definition with "inadmissible brevity, i.e., incompleteness," which "does not give us the means of deciding whether by 'equivalence' is meant a logical-formal, or a non-logical and also structurally non-describable relation." To remove this "defect" he suggests supplementing (T) in one of the two following ways:

(T')  $X$  is true if, and only if,  $p$  is true

or

(T'')  $X$  is true if, and only if,  $p$  is the case (i.e., if what  $p$  states is the case)

Then he discusses these two new "definitions," which are supposedly free from the old, formal "defect," but which turn out to be unsatisfactory for other, non-formal reasons.

This new objection seems to arise from a misunderstanding concerning the nature of sentential connectives (and thus to be somehow related to that previously discussed). The author of the objection does not seem to realize that the phrase "*if, and only if*" (in opposition to such phrases as "*are equivalent*" or "*is equivalent to*") expresses no relation between sentences at all since it does not combine names of sentences.

In general, the whole argument is based upon an obvious confusion between sentences and their names. It suffices to point out that—in contradistinction to (T)—schemata (T') and (T'') do not give any meaningful expressions if we replace in them ' $p$ ' by a sentence; for the phrases " $p$  is

*true*” and “*p is the case*” (i.e., “*what p states is the case*”) become meaningless if ‘*p*’ is replaced by a sentence, and not by the name of a sentence (cf. Section 4).<sup>26</sup>

While the author of the objection considers schema (T) “inadmissibly brief,” I am inclined, on my part, to regard schemata (T’) and (T’’) as “inadmissibly long.” And I think even that I can rigorously prove this statement on the basis of the following definition: An expression is said to be “inadmissibly long” if (i) it is meaningless, and (ii) it has been obtained from a meaningful expression by inserting superfluous words.

## 16 Redundancy of Semantic Terms—Their Possible Elimination

The objection I am going to discuss now no longer concerns the formal correctness of the definition, but is still concerned with certain formal features of the semantic conception of truth.

We have seen that this conception essentially consists in regarding the sentence “*X is true*” as equivalent to the sentence denoted by ‘*X*’ (where ‘*X*’ stands for a name of a sentence of the object-language). Consequently, the term “*true*” when occurring in a simple sentence of the form “*X is true*” can easily be eliminated, and the sentence itself, which belongs to the meta-language, can be replaced by an equivalent sentence of the object-language; and the same applies to compound sentences provided the term “*true*” occurs in them exclusively as a part of the expressions of the form “*X is true*.”

Some people have therefore urged that the term “*true*” in the semantic sense can always be eliminated, and that for this reason the semantic conception of truth is altogether sterile and useless. And since the same considerations apply to other semantic notions, the conclusion has been drawn that semantics as a whole is a purely verbal game and at best only a harmless hobby.

But the matter is not quite so simple.<sup>27</sup> The sort of elimination here discussed cannot always be made. It cannot be done in the case of universal statements which express the fact that all sentences of a certain type are true, or that all true sentences have a certain property. For instance, we can prove in the theory of truth the following statement:

All consequences of true sentences are true.

However, we cannot get rid here of the word "*true*" in the simple manner contemplated.

Again, even in the case of particular sentences having the form "*X is true*" such a simple elimination cannot always be made. In fact, the elimination is possible only in those cases in which the name of the sentence which is said to be true occurs in a form that enables us to reconstruct the sentence itself. For example, our present historical knowledge does not give us any possibility of eliminating the word "*true*" from the following sentence:

The first sentence written by Plato is true.

Of course, since we have a definition for truth and since every definition enables us to replace the definiendum by its definiens, an elimination of the term "*true*" in its semantic sense is always theoretically possible. But this would not be the kind of simple elimination discussed above, and it would not result in the replacement of a sentence in the meta-language by a sentence in the object-language.

If, however, anyone continues to urge that—because of the theoretical possibility of eliminating the word "*true*" on the basis of its definition—the concept of truth is sterile, he must accept the further conclusion that all defined notions are sterile. But this outcome is so absurd and so unsound historically that any comment on it is unnecessary. In fact, I am rather inclined to agree with those who maintain that the moments of greatest creative advancement in science frequently coincide with the introduction of new notions by means of definition.

### 17 Conformity of the Semantic Conception of Truth with Philosophical and Common-Sense Usage

The question has been raised whether the semantic conception of truth can indeed be regarded as a precise form of the old, classical conception of this notion.

Various formulations of the classical conception were quoted in the early part of this paper (Section 3). I must repeat that in my judgment none of them is quite precise and clear. Accordingly, the only sure way of settling the question would be to confront the authors of those statements with our new formulation, and to ask them whether it agrees with

their intentions. Unfortunately, this method is impractical since they died quite some time ago.

As far as my own opinion is concerned, I do not have any doubts that our formulation does conform to the intuitive content of that of Aristotle. I am less certain regarding the later formulations of the classical conception, for they are very vague indeed.<sup>28</sup>

Furthermore, some doubts have been expressed whether the semantic conception does reflect the notion of truth in its common-sense and everyday usage. I clearly realize (as I already indicated) that the common meaning of the word “*true*”—as that of any other word of everyday language—is to some extent vague, and that its usage more or less fluctuates. Hence the problem of assigning to this word a fixed and exact meaning is relatively unspecified, and every solution of this problem implies necessarily a certain deviation from the practice of everyday language.

In spite of all this, I happen to believe that the semantic conception does conform to a very considerable extent with the common-sense usage—although I readily admit I may be mistaken. What is more to the point, however, I believe that the issue raised can be settled scientifically, though of course not by a deductive procedure, but with the help of the statistical questionnaire method. As a matter of fact, such research has been carried on, and some of the results have been reported at congresses and in part published.<sup>29</sup>

I should like to emphasize that in my opinion such investigations must be conducted with the utmost care. Thus, if we ask a high-school boy, or even an adult intelligent man having no special philosophical training, whether he regards a sentence to be true if it agrees with reality, or if it designates an existing state of affairs, it may simply turn out that he does not understand the question; in consequence his response, whatever it may be, will be of no value for us. But his answer to the question whether he would admit that the sentence “*it is snowing*” could be true although it is not snowing, or could be false although it is snowing, would naturally be very significant for our problem.

Therefore, I was by no means surprised to learn (in a discussion devoted to these problems) that in a group of people who were questioned only 15% agreed that “*true*” means for them “*agreeing with reality*,” while

90% agreed that a sentence such as “*it is snowing*” is true if, and only if, it is snowing. Thus, a great majority of these people seemed to reject the classical conception of truth in its “philosophical” formulation, while accepting the same conception when formulated in plain words (waiving the question whether the use of the phrase “the same conception” is here justified).

## 18 The Definition in Its Relation to “The Philosophical Problem of Truth” and to Various Epistemological Trends

I have heard it remarked that the formal definition of truth has nothing to do with “the philosophical problem of truth.”<sup>30</sup> However, nobody has ever pointed out to me in an intelligible way just what this problem is. I have been informed in this connection that my definition, though it states necessary and sufficient conditions for a sentence to be true, does not really grasp the “essence” of this concept. Since I have never been able to understand what the “essence” of a concept is, I must be excused from discussing this point any longer.

In general, I do not believe that there is such a thing as “the philosophical problem of truth.” I do believe that there are various intelligible and interesting (but not necessarily philosophical) problems concerning the notion of truth, but I also believe that they can be exactly formulated and possibly solved only on the basis of a precise conception of this notion.

While on the one hand the definition of truth has been blamed for not being philosophical enough, on the other a series of objections have been raised charging this definition with serious philosophical implications, always of a very undesirable nature. I shall discuss now one special objection of this type; another group of such objections will be dealt with in the next section.<sup>31</sup>

It has been claimed that—due to the fact that a sentence like “snow is white” is taken to be semantically true if snow is *in fact* white (italics by the critic)—logic finds itself involved in a most uncritical realism.<sup>32</sup>

If there were an opportunity to discuss the objection with its author, I should raise two points. First, I should ask him to drop the words “*in fact*,” which do not occur in the original formulation and which are misleading, even if they do not affect the content. For these words convey

the impression that the semantic conception of truth is intended to establish the conditions under which we are warranted in asserting any given sentence, and in particular any empirical sentence. However, a moment's reflection shows that this impression is merely an illusion; and I think that the author of the objection falls victim to the illusion which he himself created.

In fact, the semantic definition of truth implies nothing regarding the conditions under which a sentence like (1):

(1) Snow is white

can be asserted. It implies only that, whenever we assert or reject this sentence, we must be ready to assert or reject the correlated sentence (2):

(2) The sentence "Snow is white" is true.

Thus, we may accept the semantic conception of truth without giving up any epistemological attitude we may have had; we may remain naive realists, critical realists or idealists, empiricists or metaphysicians—whatever we were before. The semantic conception is completely neutral toward all these issues.

In the second place, I should try to get some information regarding the conception of truth which (in the opinion of the author of the objection) does not involve logic in a most naive realism. I would gather that this conception must be incompatible with the semantic one. Thus, there must be sentences which are true in one of these conceptions without being true in the other. Assume, e.g., the sentence (1) to be of this kind. The truth of this sentence in the semantic conception is determined by an equivalence of the form (T):

The sentence "Snow is white" is true if, and only if, snow is white.

Hence in the new conception we must reject this equivalence, and consequently we must assume its denial:

The sentence "Snow is white" is true if, and only if, snow is not white (or perhaps: snow, in fact, is not white).

This sounds somewhat paradoxical. I do not regard such a consequence of the new conception as absurd; but I am a little fearful that someone in the future may charge this conception with involving logic in a "most

sophisticated kind of irrationalism.” At any rate, it seems to me important to realize that every conception of truth which is incompatible with the semantic one carries with it consequences of this type.

I have dwelt a little on this whole question, not because the objection discussed seems to me very significant, but because certain points which have arisen in the discussion should be taken into account by all those who for various epistemological reasons are inclined to reject the semantic conception of truth.

### Notes

1. Compare Tarski [2] (see bibliography at the end of the paper). This work may be consulted for a more detailed and formal presentation of the subject of the paper, especially of the material included in Sections 6 and 9–13. It contains also references to my earlier publications on the problems of semantics (a communication in Polish, 1930; the article Tarski [1] in French, 1931; a communication in German, 1932; and a book in Polish, 1933). The expository part of the present paper is related in its character to Tarski [3]. My investigations on the notion of truth and on theoretical semantics have been reviewed or discussed in Hofstadter [1], Juhos [1], Kokoszyńska [1] and [2], Kotarbiński [2], Scholz [1], Weinberg [1], *et al.*
2. It may be hoped that the interest in theoretical semantics will now increase, as a result of the recent publication of the important work Carnap [2].
3. This applies, in particular, to public discussions during the I. International Congress for the Unity of Science (Paris, 1935) and the Conference of International Congresses for the Unity of Science (Paris, 1937); cf., e.g., Neurath [1] and Gonsseth [1].
4. The words “notion” and “concept” are used in this paper with all of the vagueness and ambiguity with which they occur in philosophical literature. Thus, sometimes they refer simply to a term, sometimes to what is meant by a term, and in other cases to what is denoted by a term. Sometimes it is irrelevant which of these interpretations is meant; and in certain cases perhaps none of them applies adequately. While on principle I share the tendency to avoid these words in any exact discussion, I did not consider it necessary to do so in this informal presentation.
5. For our present purposes it is somewhat more convenient to understand by “expressions,” “sentences,” etc., not individual inscriptions, but classes of inscriptions of similar form (thus, not individual physical things, but classes of such things).
6. For the Aristotelian formulation see Aristotle [1], Γ, 7, 27. The other two formulations are very common in the literature, but I do not know with whom they originate. A critical discussion of various conceptions of truth can be found,



e.g., in Kotarbiński [1] (so far available only in Polish), pp. 123 ff., and Russell [1], pp. 362 ff.

7. For most of the remarks contained in Sections 4 and 8, I am indebted to the late S. Leśniewski who developed them in his unpublished lectures in the University of Warsaw (in 1919 and later). However, Leśniewski did not anticipate the possibility of a rigorous development of the theory of truth, and still less of a definition of this notion; hence, while indicating equivalences of the form (T) as premisses in the antinomy of the liar, he did not conceive them as any sufficient conditions for an adequate usage (or definition) of the notion of truth. Also the remarks in Section 8 regarding the occurrence of an empirical premiss in the antinomy of the liar, and the possibility of eliminating this premiss, do not originate with him.

8. In connection with various logical and methodological problems involved in this paper the reader may consult Tarski [6].

9. The antinomy of the liar (ascribed to Eubulides or Epimenides) is discussed here in Sections 7 and 8. For the antinomy of definability (due to J. Richard) see, e.g., Hilbert-Bernays [1], vol. 2, pp. 263 ff.; for the antinomy of heterological terms see Grelling-Nelson [1], p. 307.

10. Due to Professor J. Łukasiewicz (University of Warsaw).

11. This can roughly be done in the following way. Let *S* be any sentence beginning with the words “*Every sentence.*” We correlate with *S* a new sentence *S\** by subjecting *S* to the following two modifications: we replace in *S* the first word, “*Every,*” by “*The*”; and we insert after the second word, “*sentence,*” the whole sentence *S* enclosed in quotation marks. Let us agree to call the sentence *S* “(self-)applicable” or “non-(self-)applicable” dependent on whether the correlated sentence *S\** is true or false. Now consider the following sentence:

Every sentence is non-applicable.

It can easily be shown that the sentence just stated must be both applicable and non-applicable; hence a contradiction. It may not be quite clear in what sense this formulation of the antinomy does not involve an empirical premiss; however, I shall not elaborate on this point.

12. The terms “logic” and “logical” are used in this paper in a broad sense, which has become almost traditional in the last decades; logic is assumed here to comprehend the whole theory of classes and relations (i.e., the mathematical theory of sets). For many different reasons I am personally inclined to use the term “logic” in a much narrower sense, so as to apply it only to what is sometimes called “elementary logic,” i.e., to the sentential calculus and the (restricted) predicate calculus.

13. Cf. here, however, Tarski [3], pp. 5 f.

14. The method of construction we are going to outline can be applied—with appropriate changes—to all formalized languages that are known at the present time; although it does not follow that a language could not be constructed to which this method would not apply.

15. In carrying through this idea a certain technical difficulty arises. A sentential function may contain an arbitrary number of free variables; and the logical nature of the notion of satisfaction varies with this number. Thus, the notion in question when applied to functions with one variable is a binary relation between these functions and single objects; when applied to functions with two variables it becomes a ternary relation between functions and couples of objects; and so on. Hence, strictly speaking, we are confronted, not with one notion of satisfaction, but with infinitely many notions; and it turns out that these notions cannot be defined independently of each other, but must all be introduced simultaneously.

To overcome this difficulty, we employ the mathematical notion of an infinite sequence (or, possibly, of a finite sequence with an arbitrary number of terms). We agree to regard satisfaction, not as a many-termed relation between sentential functions and an indefinite number of objects, but as a binary relation between functions and sequences of objects. Under this assumption the formulation of a general and precise definition of satisfaction no longer presents any difficulty; and a true sentence can now be defined as one which is satisfied by every sequence.

16. To define recursively the notion of satisfaction, we have to apply a certain form of recursive definition which is not admitted in the object-language. Hence the "essential richness" of the meta-language may simply consist in admitting this type of definition. On the other hand, a general method is known which makes it possible to eliminate all recursive definitions and to replace them by normal, explicit ones. If we try to apply this method to the definition of satisfaction, we see that we have either to introduce into the meta-language variables of a higher logical type than those which occur in the object-language; or else to assume axiomatically in the meta-language the existence of classes that are more comprehensive than all those whose existence can be established in the object-language. See here Tarski [2], pp. 393 ff., and Tarski [5], p. 110.

17. Due to the development of modern logic, the notion of mathematical proof has undergone a far-reaching simplification. A sentence of a given formalized discipline is provable if it can be obtained from the axioms of this discipline by applying certain simple and purely formal rules of inference, such as those of detachment and substitution. Hence to show that all provable sentences are true, it suffices to prove that all the sentences accepted as axioms are true, and that the rules of inference when applied to true sentences yield new true sentences; and this usually presents no difficulty.

On the other hand, in view of the elementary nature of the notion of provability, a precise definition of this notion requires only rather simple logical devices. In most cases, those logical devices which are available in the formalized discipline itself (to which the notion of provability is related) are more than sufficient for this purpose. We know, however, that as regards the definition of truth just the opposite holds. Hence, as a rule, the notions of truth and provability cannot coincide; and since every provable sentence is true, there must be true sentences which are not provable.

18. Thus the theory of truth provides us with a general method for consistency proofs for formalized mathematical disciplines. It can be easily realized, however,

that a consistency proof obtained by this method may possess some intuitive value—i.e., may convince us, or strengthen our belief, that the discipline under consideration is actually consistent—only in case we succeed in defining truth in terms of a meta-language which does not contain the object-language as a part (cf. here a remark in Section 9). For only in this case the deductive assumptions of the meta-language may be intuitively simpler and more obvious than those of the object-language—even though the condition of “essential richness” will be formally satisfied. Cf. here also Tarski [3], p. 7.

The incompleteness of a comprehensive class of formalized disciplines constitutes the essential content of a fundamental theorem of K. Gödel; cf. Gödel [1], pp. 187 ff. The explanation of the fact that the theory of truth leads so directly to Gödel’s theorem is rather simple. In deriving Gödel’s result from the theory of truth we make an essential use of the fact that the definition of truth cannot be given in a meta-language which is only as “rich” as the object-language (cf. note 17); however, in establishing this fact, a method of reasoning has been applied which is very closely related to that used (for the first time) by Gödel. It may be added that Gödel was clearly guided in his proof by certain intuitive considerations regarding the notion of truth, although this notion does not occur in the proof explicitly; cf. Gödel [1], pp. 174 f.

19. The notions of designation and definition lead respectively to the antinomies of Grelling-Nelson and Richard (cf. note 9). To obtain an antinomy for the notion of satisfaction, we construct the following expression:

The sentential function *X* does not satisfy *X*.

A contradiction arises when we consider the question whether this expression, which is clearly a sentential function, satisfies itself or not.

20. All notions mentioned in this section can be defined in terms of satisfaction. We can say, e.g., that a given term designates a given object if this object satisfies the sentential function “*x is identical with T*” where ‘*T*’ stands for the given term. Similarly, a sentential function is said to define a given object if the latter is the only object which satisfies this function. For a definition of consequence see Tarski [4], and for that of synonymy—Carnap [2].

21. General semantics is the subject of Carnap [2]. Cf. here also remarks in Tarski [2], pp. 388 f.

22. Cf. various quotations in Ness [1], pp. 13 f.

23. The names of persons who have raised objections will not be quoted here, unless their objections have appeared in print.

24. It should be emphasized, however, that as regards the question of an alleged vicious circle the situation would not change even if we took a different point of view, represented, e.g., in Carnap [2]; i.e., if we regarded the specification of conditions under which sentences of a language are true as an essential part of the description of this language. On the other hand, it may be noticed that the point of view represented in the text does not exclude the possibility of using truth-

tables in a deductive development of logic. However, these tables are to be regarded then merely as a formal instrument for checking the provability of certain sentences; and the symbols 'T' and 'F' which occur in them and which are usually considered abbreviations of "true" and "false" should not be interpreted in any intuitive way.

25. Cf. Juhos [1]. I must admit that I do not clearly understand von Juhos' objections and do not know how to classify them; therefore, I confine myself here to certain points of a formal character. Von Juhos does not seem to know my definition of truth; he refers only to an informal presentation in Tarski [3] where the definition has not been given at all. If he knew the actual definition, he would have to change his argument. However, I have no doubt that he would discover in this definition some "defects" as well. For he believes he has proved that "on ground of principle it is impossible to give such a definition at all."

26. The phrases "*p is true*" and "*p is the case*" (or better "*it is true that p*" and "*it is the case that p*") are sometimes used in informal discussions, mainly for stylistic reasons; but they are considered then as synonymous with the sentence represented by '*p*'. On the other hand, as far as I understand the situation, the phrases in question cannot be used by von Juhos synonymously with '*p*'; for otherwise the replacement of (T) by (T') or (T'') would not constitute any "improvement."

27. Cf. the discussion of this problem in Kokoszyńska [1], pp. 161 ff.

28. Most authors who have discussed my work on the notion of truth are of the opinion that my definition does conform with the classical conception of this notion; see, e.g., Kotarbiński [2] and Scholz [1].

29. Cf. Ness [1]. Unfortunately, the results of that part of Ness' research which is especially relevant for our problem are not discussed in his book; compare p. 148, footnote 1.

30. Though I have heard this opinion several times, I have seen it in print only once and, curiously enough, in a work which does not have a philosophical character—in fact, in Hilbert-Bernays [1], vol. II, p. 269 (where, by the way, it is not expressed as any kind of objection). On the other hand, I have not found any remark to this effect in discussions of my work by professional philosophers (cf. note 1).

31. The remaining sections of Tarski's original article have been omitted in this reprinting.—Ed.

32. Cf. Gonsseth [1], pp. 187 f.

## Bibliography

Only the books and articles actually referred to in the paper will be listed here.

Aristotle [1]. *Metaphysics*. (*Works*, vol. VIII.) English translation by W. D. Ross. Oxford, 1908.

- Carnap, R. [1]. *Logical Syntax of Language*. London and New York, 1937.
- Carnap, R. [2]. *Introduction to Semantics*. Cambridge, 1942.
- Gödel, K. [1]. "Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme, I." *Monatshefte für Mathematik und Physik*, vol. xxxviii, 1931, pp. 173–198.
- Gödel, K. [2]. "Über die Länge von Beweisen." *Ergebnisse eines mathematischen Kolloquiums*, vol. vii, 1936, pp. 23–24.
- Gonseth, F. [1]. "Le Congrès Descartes. Questions de Philosophie scientifique." *Revue thomiste*, vol. xlv, 1938, pp. 183–193.
- Grelling, K., and Nelson, L. [1]. "Bemerkungen zu den Paradoxien von Russell und Burali-Forti." *Abhandlungen der Fries'schen Schule*, vol. ii (new series), 1908, pp. 301–334.
- Hofstadter, A. [1]. "On Semantic Problems." *Journal of Philosophy*, vol. xxxv, 1938, pp. 225–232.
- Hilbert, D., and Bernays, P. [1]. *Grundlagen der Mathematik*. 2 vols. Berlin, 1934–1939.
- Juhos, B. von. [1]. "The Truth of Empirical Statements." *Analysis*, vol. iv, 1937, pp. 65–70.
- Kokoszyńska, M. [1]. "Über den absoluten Wahrheitsbegriff und einige andere semantische Begriffe." *Erkenntnis*, vol. vi, 1936, pp. 143–165.
- Kokoszyńska, M. [2]. "Syntax, Semantik und Wissenschaftslogik." *Actes du Congrès International de Philosophie Scientifique*, vol. iii, Paris, 1936, pp. 9–14.
- Kotarbiński, T. [1]. *Elementy teorii poznania, logiki formalnej i metodologii nauk*. (Elements of Epistemology, Formal Logic, and the Methodology of Sciences, in Polish.) Lwów, 1929.
- Kotarbiński, T. [2]. "W sprawie pojęcia prawdy." ("Concerning the Concept of Truth," in Polish.) *Przegląd filozoficzny*, vol. xxxvii, pp. 85–91.
- Lindenbaum, A., and Tarski, A. [1]. "Über die Beschränktheit der Ausdrucksmittel deduktiver Theorien." *Ergebnisse eines mathematischen Kolloquiums*, vol. vii, 1936, pp. 15–23.
- Nagel, E. [1]. Review of Hofstadter [1]. *Journal of Symbolic Logic*, vol. iii, 1938, p. 90.
- Nagel, E. [2]. Review of Carnap [2]. *Journal of Philosophy*, vol. xxxix, 1942, pp. 468–473.
- Ness, A. [1]. "'Truth' As Conceived by Those Who Are Not Professional Philosophers." *Skrifter utgitt av Det Norske Videnskaps-Akademi i Oslo, II. Hist.-Filos. Klasse*, vol. iv, Oslo, 1938.
- Neurath, O. [1]. "Erster Internationaler Kongress für Einheit der Wissenschaft in Paris 1935." *Erkenntnis*, vol. v, 1935, pp. 377–406.
- Russell, B. [1]. *An Inquiry Into Meaning and Truth*. New York, 1940.

Scholz, H. [1]. Review of *Studia philosophica*, vol. i. *Deutsche Literaturzeitung*, vol. LVIII, 1937, pp. 1914–1917.

Tarski, A. [1]. “Sur les ensembles définissables de nombres réels, I.” *Fundamenta mathematicae*, vol. XVII, 1931, pp. 210–239.

Tarski, A. [2]. “Der Wahrheitsbegriff in den formalisierten Sprachen.” (German translation of a book in Polish, 1933.) *Studia philosophica*, vol. i, 1935, pp. 261–405.

Tarski, A. [3]. “Grundlegung der wissenschaftlichen Semantik.” *Actes du Congrès International de Philosophie Scientifique*, vol. III, Paris, 1936, pp. 1–8.

Tarski, A. [4]. “Über den Begriff der logischen Folgerung.” *Actes du Congrès International de Philosophie Scientifique*, vol. VII, Paris, 1937, pp. 1–11.

Tarski, A. [5]. “On Undecidable Statements in Enlarged Systems of Logic and the Concept of Truth.” *Journal of Symbolic Logic*, vol. IV, 1939, pp. 105–112.

Tarski, A. [6]. *Introduction to Logic*. New York, 1941.

Weinberg, J. [1]. Review of *Studia philosophica*, vol. i. *The Philosophical Review*, vol. XLVII, pp. 70–77.



---

## Tarski's Theory of Truth

Hartry Field

In the early 1930s there was prevalent, among scientifically minded philosophers, the view that semantic notions such as the notions of truth and denotation were illegitimate: that they could not or should not be incorporated into a scientific conception of the world. But when Tarski's work on truth became known, all this changed. Popper wrote, "As a result of Tarski's teaching, I no longer hesitate to speak of 'truth' and 'falsity'";<sup>1</sup> and Popper's reaction was widely shared.<sup>2</sup>

A philosopher who shared Popper's reaction to Tarski's discoveries would presumably argue as follows. "What Tarski did was to define the term 'true', using in his definitions only terms that are clearly acceptable. In particular, he did not employ any undefined semantic terms in his definitions. So Tarski's work should make the term 'true' acceptable even to someone who is initially suspicious of semantic terms."

This contention has an initial plausibility, but I will argue that it is radically wrong. My contrary claim will be that Tarski succeeded in reducing the notion of truth *to certain other semantic notions*; but that he did not in any way explicate these other notions, so that his results ought to make the word 'true' acceptable only to someone who already regarded these other semantic notions as acceptable.

By claiming that Tarski merely reduced truth to other semantic notions, I don't mean to suggest that his results on truth are trivial. On the contrary, I think that they are extremely important, and have applications not only to mathematics but also to linguistics and to more directly philosophical problems about realism and objectivity. I think, however, that the real value of Tarski's discoveries for linguistics and philosophy is widely misunderstood, and I hope to eradicate the most central mis-



understandings by clarifying and defending the claim that Tarski merely reduced truth to other semantic notions.

## I

I believe that Tarski presented his semantic theory in a very misleading way, one which has encouraged the misinterpretations just alluded to. In this section I will present Tarski's theory as I think he should have presented it. However, I do not expect instant agreement that this new way is better than the old, and so I will use the name 'Tarski\*' for a logician who gave the sort of semantic theory I will now sketch. Later in the paper I will compare Tarski\*'s semantics to the semantics that the real Tarski actually gave; by doing this I will cast light on the issues raised in my introductory paragraphs.

In sketching Tarski\*'s theory, I will focus my attention on a particular object language *L*. The language *L* that I choose will be a quantificational language with names ('*c*<sub>1</sub>', '*c*<sub>2</sub>', ...), one-place function symbols ('*f*<sub>1</sub>', '*f*<sub>2</sub>', ...), and one-place predicates ('*p*<sub>1</sub>', '*p*<sub>2</sub>', ...). The language of course cannot be viewed as an "uninterpreted" language, i.e., as just a bunch of strings of meaningless marks, for then there would be no truth to worry about. Instead, the language should be regarded as something that people actually speak or write; and it is because the speakers speak or write the way they do that the words of the language have the meaning they have.<sup>3</sup>

Initially I will follow Tarski in supposing that in *L* "the sense of every expression is unambiguously determined by its form,"<sup>4</sup> i.e., that whenever two speakers use the same name (or one speaker uses it on two occasions) they are referring to the same thing, that whenever two speakers use the same sentence either both are saying something true or neither is, etc. In these circumstances it makes sense to speak of the names of the language denoting things (a name denotes whatever the users of the name refer to) and the sentences being true or false (true when speakers who use it say something true by so doing.) The more general situation, in which there are expressions whose "sense" is not determined wholly by their form, will be dealt with later. (We'll see that it is one of the advantages of Tarski\*'s semantics that it can easily handle this more general situation).

The syntax of  $L$  can be given by two recursive definitions: first we define the *singular terms* by saying that all names and variables are singular terms, and a function symbol followed by a singular term is a singular term; then we define the *formulas* by saying that a predicate followed by a singular term is a formula, as is the negation of a formula, the conjunction of two formulas, and the universal quantification of a formula with any variable. The *sentences*, or *closed formulas*, are then singled out in the usual way.

Now we can proceed to Tarski's semantics. Rather than characterize truth directly, we characterize it relative to some assignment of objects to the variables, say  $s_k$  to ' $x_k$ '. The idea is going to be to treat the variables, or at least the free variables, as sort of "temporary names" for the objects assigned to them. So we proceed by fixing a sequence  $s = \langle s_1, s_2, \dots \rangle$  of objects, to be assigned to ' $x_1$ ', ' $x_2$ ',  $\dots$ , respectively; and we want to say what it is for a formula to be true <sub>$s$</sub> , i.e., true relative to the assignment  $s$ . As a preliminary we say what it is for a term to denote <sub>$s$</sub>  an object, i.e., to denote it relative to the assignment  $s$ . The denotation of ' $x_k$ ' relative to  $s$  is evidently  $s_k$ , for this is the object assigned to ' $x_k$ '. But what is the denotation relative to  $s$  of ' $c_k$ '? Evidently what objects are assigned to the variables here is irrelevant, and the denotation <sub>$s$</sub>  of ' $c_k$ ' is some fixed object that users of the language refer to when they use the name ' $c_k$ '. Just what this object is depends on facts we have not yet been given about the use of ' $c_k$ '. Similarly there are facts we have not yet been given about the use of ' $p_k$ ' and ' $f_k$ ' which we need in order to fix the truth value of sentences containing them. For ' $p_k$ ' the relevant facts concern the extension of the predicate—what objects the predicate *applies to*—for it is this which affects the truth value of all utterances containing ' $p_k$ '. For ' $f_k$ ', the relevant facts concern what pairs of objects *fulfill* that function symbol—in the sense that the pair  $\langle \text{John Adams}, \text{John Quincy Adams} \rangle$  and every other father-son pair fulfill the function symbol 'father of'.

With these points in mind it is now easy to give an inductive characterization of denotation <sub>$s$</sub> :

- T1 (A) 1. ' $x_k$ ' denotes <sub>$s$</sub>   $s_k$ .  
 2. ' $c_k$ ' denotes <sub>$s$</sub>  what it denotes.  
 3. ' $f_k(e)$ ' denotes <sub>$s$</sub>  an object  $a$  if and only if

- (i) there is an object  $b$  that  $e$  denotes<sub>s</sub> and
- (ii) ' $f_k$ ' is fulfilled by  $\langle a, b \rangle$ .

(Here ' $e$ ' is a variable ranging over expressions of L.) Similarly we define 'true<sub>s</sub>' for formulas—what Tarski calls satisfaction of a formula by  $s$ :

- (B) 1.  $\lceil p_k(e) \rceil$  is true<sub>s</sub> if and only if
- (i) there is an object  $a$  that  $e$  denotes<sub>s</sub> and
  - (ii) ' $p_k$ ' applies to  $a$ .
2.  $\lceil \sim e \rceil$  is true<sub>s</sub> if and only if  $e$  is not true<sub>s</sub>.
3.  $\lceil e_1 \wedge e_2 \rceil$  is true<sub>s</sub> if and only if  $e_1$  is true<sub>s</sub> and so is  $e_2$ .
4.  $\lceil \forall x_k(e) \rceil$  is true<sub>s</sub> if and only if for each sequence  $s^*$  that differs from  $s$  at the  $k$ th place at most,  $e$  is true<sub>s\*</sub>.

This completes the characterization of truth relative to an assignment of objects to the variables. In the case of sentences it is easily seen that we get the same results whatever such assignment we pick; we can say

- (C) A sentence is true if and only if its is true<sub>s</sub> for some (or all)  $s$ .

This completes my elaboration of Tarski's "truth definition" T1 for L—or his *truth characterization* (TC), as I prefer to call it. What is its philosophical significance? The obvious answer, and the correct one, I think, is that the TC reduces one semantic notion to three others. It explains what it is for a sentence to be true in terms of certain semantic features of the primitive components of the sentence: in terms of what it is for a name to denote something, what it is for a predicate to apply to something, and what it is for a function symbol to be fulfilled by some pair of things. It is convenient to introduce the expression 'primitively denotes' as follows: every name *primitively denotes* what it denotes; every predicate and every function symbol *primitively denotes* what it applies to or is fulfilled by; and no complex expression primitively denotes anything. In this terminology, what T1 does is to explain truth in terms of primitive denotation. Similarly we can explain denotation for arbitrary closed singular terms [such as ' $f_1(c_1)$ '] in terms of primitive denotation, i.e., in terms of the semantic features of the names and function symbols from which the complex singular term is composed—we merely say that a closed singular term denotes an object  $a$  if it denotes<sub>s</sub>  $a$  for some (or all)  $s$ , where denotation<sub>s</sub> is defined as before. We see then that *Tarski's semantics explains the semantic properties of complex expres-*

sions (e.g., truth value for sentences, denotation for complex singular terms) *in terms of semantic properties of their primitive components*.

To explain truth in terms of primitive denotation is, I think, an important task. It certainly doesn't answer *every* question that anyone would ever want answered about truth, but for many purposes it is precisely what we need. For instance, in model theory we are interested in such questions as: given a set  $\Gamma$  of sentences, is there any way to choose the denotations of the primitives of the language so that every sentence of  $\Gamma$  will come out true given the usual semantics for the logical connectives?<sup>5</sup> For questions such as this, what we need to know is how the truth value of a whole sentence depends on the denotations of its primitive nonlogical parts, and that is precisely what T1 tells us. So *at least for model-theoretic purposes*, Tarski's TC is precisely the kind of explication of truth we need.

I want now to return to a point I mentioned earlier, about Tarski's restriction to languages in which "the sense of every expression is unambiguously determined by its form." Natural languages are full of expressions that do not meet this requirement. For instance, different tokens of 'John takes grass' can differ in "sense"—e.g., one token may be uttered in saying that John Smith smokes marijuana, and another may be uttered in saying that John Jones steals lawn material, and these differences may give rise to differences of truth value in the tokens. (I say that a complete<sup>6</sup> token of a sentence is true if the person who spoke or wrote that token said something true by so doing; I also say that a name token denotes an object if the person who spoke or wrote the token referred to the object by so doing.) The prevalence of such examples in natural languages raises the question of whether Tarski's type of semantic theory is applicable to languages in which the sense is *not* determined by the form; for if the answer is no, then Davidson's very worth-while project<sup>7</sup> of giving truth characterizations for natural languages seems doomed from the start.

It seems clear that if we stick to the kind of TC that Tarski actually gave (see next section), there is no remotely palatable way of extending TC's to sentences like 'John takes grass'. But if we use TC's like T1 there is no difficulty at all. The only point about languages containing 'John' or 'grass' or 'I' or 'you' is that for such languages 'true', 'denotes', and other

semantic terms make no clear sense as applied to expression types; they make sense only as applied to tokens. For this reason we have to interpret clause (B)2 of T1 as meaning

A token of  $\ulcorner \sim e \urcorner$  is true<sub>s</sub> if and only if the token of  $e$  that it contains is not true<sub>s</sub>.

and similarly for the other clauses. Once we interpret out TC in this way in terms of tokens, i.e., individual occasions of utterance, that TC works perfectly: someone who utters 'John is sick' (or 'I am sick') says something true if and only if his token of 'sick' applies to the person he refers to by 'John' (or by 'I'); and the fact that other speakers (or this speaker on other occasions) sometimes refer to different things when they use 'John' (or 'I') is beside the point.

This analysis leaves entirely out of account the ways in which 'I' and 'John' differ: it leaves out of account, for instance, the fact that a token of 'I' always denotes the speaker who produced it. But that is no objection to the analysis, for the analysis purports merely to explain truth in terms of primitive denotation; it does not purport to say anything about primitive denotation, and the differences between 'I' and 'John' (or their analogues in a language like L) are purely differences of how they denote. (The word 'I' denotes according to the simple rule mentioned two sentences back; 'John' denotes according to much more complex rules that I have no idea how to formulate.)

Of course, the fact that a theory of denotation for a word like 'I' is so simple and obvious, makes it possible to alter the TC so that the theory of denotation for such a word is built into the TC itself—such a course is adopted, for instance, by Davidson at the end of "Truth and Meaning." I myself prefer to preserve the analogies of the word 'I' to words that function less systematically, e.g., 'we', 'she', and 'John'. How one treats 'I' is more or less a matter of taste; but the less systematic words I've just mentioned cannot be handled in the way that Davidson handles 'I', and the only reasonable way I can see to handle them is the way I have suggested: use a truth characterization like T1 (except stated in terms of tokens rather than types), and leave it to a separate theory of primitive denotation to explain the relevant differences between tokens of 'John' that denote John Adams and tokens of 'John' that denote John Lennon,

and between tokens of 'bank' that apply to things along rivers and tokens of 'bank' that apply to the Chase Manhattan.<sup>8</sup>

There are other advantages to T1 besides its ability to handle ambiguous sentences, i.e., sentences for which the sense is not determined by the form. For instance, Tarski required that the vocabulary of the language be fixed once and for all; but if we decide to give truth characterizations of type T1, this is unnecessary: all that is required is that the general structure of the language be fixed, e.g., that the semantic categories<sup>9</sup> (name, one-place predicate, etc.) be held constant. In other words, if a language already contained proper names, the invention of a new name to baptize an object will not invalidate the old TC; though introduction of a name into a hitherto nameless language will.

To show this, we have merely to reformulate the given TC so that it does not rely on the actual vocabulary that the language contains at a given time, but works also for sentences containing new names, one-place predicates, etc., that speakers of the language might later introduce. To do this is trivial: we define denotation<sub>s</sub> by

1. The  $k$ th variable denotes<sub>s</sub>  $s_k$ .
2. If  $e_1$  is a name, it denotes<sub>s</sub> what it denotes.
3. If  $e_1$  is a singular term and  $e_2$  is a function symbol, then  $\ulcorner e_2(e_1) \urcorner$  denotes<sub>s</sub>  $a$  if and only if
  - (i) as before, and
  - (ii)  $e_2$  is fulfilled by  $\langle a, b \rangle$ .

And we can generalize the definition of truth<sub>s</sub> in a similar manner.<sup>10</sup> This shows that, in giving a TC, there is no need to utilize the particular vocabulary used at one temporal stage of a language, for we can instead give a more general TC which can be incorporated into a diachronic theory of the language (and can also be applied directly to other languages of a similar structure). If, that is, we accept the modification of Tarski proposed in this section.

## II

The kind of truth characterization advocated in the previous section differs from the kind of TC Tarski offered in one important respect. Tarski

stated the policy “I shall not make use of any semantical concept if I am not able previously to reduce it to other concepts” (CTFL 152/3), and this policy is flagrantly violated by T1: T1 utilizes unreduced notions of proper names denoting things, predicates applying to things, and function symbols being fulfilled by things.

Tarski’s truth characterizations, unlike T1, accorded with his stated policy: they did not contain any semantic terms like ‘applies to’ or ‘denotes’. How did Tarski achieve this result? Very simply: first, he translated every name, predicate, and function symbol of *L* into English; then he utilized these translations in order to reformulate clauses 2 and 3(ii) of part (A) of the definition and clause 1(ii) of part (B). For simplicity, let’s use ‘ $\bar{c}_1$ ’, ‘ $\bar{c}_2$ ’, etc. as abbreviations for the English expressions that are the translations of the words ‘ $c_1$ ’, ‘ $c_2$ ’, ... of *L*: e.g.: if *L* is simplified German and ‘ $c_1$ ’ is ‘Deutschland’, then ‘ $\bar{c}_1$ ’ is an abbreviation for ‘Germany’. Similarly, let ‘ $\bar{f}_1$ ’ abbreviate the translation into English of the word ‘ $f_1$ ’ of *L*, and let ‘ $\bar{p}_1$ ’ abbreviate the translation of ‘ $p_1$ ’ into English. Then Tarski’s reformulated truth definition will read as follows:

- T2 (A) 1. as before  
       2. ‘ $c_k$ ’ denotes<sub>s</sub>  $\bar{c}_k$   
       3.  $\lceil f_k(e) \rceil$  denotes<sub>s</sub>  $a$  if and only if  
           (i) as before  
           (ii)  $a$  is  $\bar{f}_k(b)$   
 (B) 1.  $\lceil p_k(e) \rceil$  is true<sub>s</sub> if and only if  
       (i) as before  
       (ii)  $\bar{p}_k(a)$   
       2–4. as before  
 (C) as before

What T2 is like depends of course on the precise character of the translations of the primitives that are utilized. For instance, if we translate ‘ $c_1$ ’ as ‘the denotation of ‘ $c_1$ ’’, translate ‘ $p_1$ ’ as ‘is something that ‘ $p_1$ ’ applies to’, etc., then T2 becomes identical with T1. This of course is *not* what Tarski intended. What Tarski intended is that T2 not contain unexplicated semantic terms, and if we are to get this result we must not employ any semantic terms in our translations.<sup>11</sup>

But other restrictions on translations are also necessary: if we were to translate 'Deutschland' as 'Bertrand Russell', a truth characterization T2 that was based on this translation would grossly misrepresent L. In order to state the matter more generally, I introduce the term 'coreferential': two singular terms are coreferential if they denote the same thing; two predicative expressions are coreferential if they have the same extension, i.e., if they apply to the same things; and two functional expressions are coreferential if they are fulfilled by the same pairs. It is then easily seen that any departure from coreferentiality in translation will bring errors into T2. For instance, suppose we translate the foreign predicate 'glub' as 'yellow', and suppose 'glub' and yellow are not *precisely* coreferential; then clause (B)<sub>1</sub> will say falsely that 'glub(*x*)' is true of just those objects which are yellow.

Let us say, then, that

- (1) An adequate translation of a primitive  $e_1$  of  $L$  into English is an expression  $e_2$  of English such that
  - (i)  $e_1$  and  $e_2$  are coreferential, and
  - (ii)  $e_2$  contains no semantic terms.

This notion of an adequate translation is of course a semantic notion that Tarski did not reduce to nonsemantic terms. But that is no objection to his characterization T2 (at least, it isn't obviously an objection), for the notion of an adequate translation is never built into the truth characterization and is not, properly speaking, part of a theory of truth. On Tarski's view we need to adequately translate the object language into the meta-language in order to give an adequate theory of truth for the object language; this means that the notion of an adequate translation is employed in the methodology of giving truth theories, but it is not employed in the truth theories themselves.

In what follows I shall assume that the language  $L$  with which we are dealing is so related to English that all its primitives *can* be adequately translated into English, according to the standards of adequacy set forth in (1). (This is another restriction that we avoid if we give TC's of the type T1; quite a significant restriction, I think.) If we then suppose that the translation given ( $\bar{c}_1$  for ' $c_1$ ', etc.) is one of the adequate translations, then T2, like T1, is a correct recursive characterization of truth



for the language  $L$ . There is, of course, a simple procedure for transforming recursive characterizations such as these into explicit characterizations. To carry the procedure through in these cases would be pretty complicated, but it could be done; so we could regard T1 (or T2) as implicitly specifying a metalinguistic formula ' $A_1(e)$ ' (or ' $A_2(e)$ '), and saying that an utterance  $e$  of  $L$  is true if and only if  $A_1(e)$  (or  $A_2(e)$ ). If we regard T1 and T2 as written in this form, then the key difference between them is that ' $A_1(e)$ ' contains semantic terms and ' $A_2(e)$ ' does not. The question then arises: is the fact that ' $A_2(e)$ ' does not contain semantic terms an advantage of T2 over T1? If so, then *why* is it an advantage?

In order to discuss the possible advantages of T2 over T1, I think we have to go beyond mathematical considerations and focus instead on linguistic and other "philosophical" matters. It is not enough to say that T2 *defines* truth without utilizing semantic terms, whereas T1 defines it only in other semantic terms: this is not enough until we say something more about the purpose of definition. If the purpose of giving a "definition" of truth is to enable you to do model theory, then the elimination of semantic terms from T1 gives no advantage. For what purpose do we want definitions for which the elimination of semantic terms is useful?

One purpose to which definitions are sometimes put is in explaining the meaning of a word. This of course is very vague, but I think it is clear enough to enable use to recognize that neither T1 nor T2 has very much to do with explaining the meaning of the word 'true'. This is especially obvious for T2: a T2-type truth definition works for a single language only, and so if it "explains the meaning of" the word 'true' as applied to that language, then for *any* two languages  $L_1$  and  $L_2$ , the word 'true' means something different when applied to utterances of  $L_1$  than it means when applied to utterances of  $L_2$ ! I make this point not in criticism of T2, but in criticism of the idea that the significance of T2 can be explained by saying that it "gives the meaning of" the word 'true'.

We still need to know what purpose a truth characterization like T1 or T2 could serve that would give someone reason to think that a TC without unexplicated semantic terms would be better than a TC with unexplicated semantic terms. Tarski hints at such a purpose in one place in his writings, where he is discussing the importance of being able to define the word 'true', as opposed to merely introducing axioms to estab-

lish the basic properties of truth. If a definition of semantic notions such as truth could not be given, Tarski writes,

... it would then be difficult to bring [semantics] into harmony with the postulates of the unity of science and of physicalism (since the concepts of semantics would be neither logical nor physical concepts).<sup>12</sup>

This remark seems to me to be of utmost importance in evaluating the philosophical significance of Tarski's work, and so I will now say something about the general philosophical issues it raises. When this is done we will be in a better position to understand Tarski's choice of T2 over T1.

### III

In the early 1930s many philosophers believed that the notion of truth could not be incorporated into a scientific conception of the world. I think that the main rationale for this view is hinted at in the remark of Tarski's that I quoted at the end of the last section, and what I want to do now is to elaborate a bit on Tarski's hint.

In the remark I have quoted, Tarski put a heavy stress on the doctrine of physicalism: the doctrine that chemical facts, biological facts, psychological facts, and semantical facts, are all explicable (in principle) in terms of physical facts. The doctrine of physicalism functions as a high-level empirical hypothesis, a hypothesis that no small number of experiments can force us to give up. It functions, in other words, in much the same way as the doctrine of mechanism (that all facts are explicable in terms of *mechanical* facts) once functioned: this latter doctrine has now been universally rejected, but it was given up only by the development of a well-accepted theory (Maxwell's) which described phenomena (electromagnetic radiation and the electromagnetic field) that were very difficult to account for mechanically, and by amassing a great deal of experiment and theory that together made it quite conclusive that mechanical explanations of these phenomena (e.g., by positing "the ether") would never get off the ground. Mechanism has been empirically refuted; its heir is physicalism, which allows as "basic" not only facts about mechanics, but facts about other branches of physics as well.<sup>13</sup> I believe that physicists a

hundred years ago were justified in accepting mechanism, and that, similarly, physicalism should be accepted until we have convincing evidence that there is a realm of phenomena it leaves out of account. Even if there *does* turn out to be such a realm of phenomena, the only way we'll ever come to know that there is, is by repeated efforts and repeated failures to explain these phenomena in physical terms.

That's my view, anyway, but there are philosophers who think that it is in order to reject physicalism now. One way of rejecting physicalism is called "vitalism": it is the view that there are irreducibly biological facts, i.e., biological facts that aren't explicable in nonbiological terms (and hence, not in physical terms). Physicalism and vitalism are incompatible, and it is because of this incompatibility that the doctrine of physicalism has the methodological importance it has for biology. Suppose, for instance, that a certain woman has two sons, one hemophilic and one not. Then, according to standard genetic accounts of hemophilia, the ovum from which one of these sons was produced must have contained a gene for hemophilia, and the ovum from which the other son was produced must not have contained such a gene. But now the doctrine of physicalism tells us that there must have been a *physical* difference between the two ova that explains why the first son had hemophilia and the second one didn't, if the standard genetic account is to be accepted. We should not rest content with a special biological predicate 'has-a-hemophilic-gene'—rather, we should look for nonbiological facts (chemical facts; and ultimately, physical facts) that underlie the correct application of this predicate. That at least is what the principle of physicalism tells us, and it can hardly be doubted that this principle has motivated a great deal of very profitable research into the chemical foundations of genetics.

So much for vitalism; now let us turn to other irreducibility doctrines that are opposed to physicalism. One such irreducibility doctrine is Cartesianism: it is the doctrine that there are irreducibly mental facts. Another irreducibility doctrine has received much less attention than either vitalism or Cartesianism, but it is central to our present concerns: this doctrine, which might be called "semanticalism," is the doctrine that there are irreducibly semantic facts. The semanticalist claims, in other words, that semantic phenomena (such as the fact that 'Schnee' refers to snow) must be accepted as primitive, in precisely the way that electro-

magnetic phenomena are accepted as primitive (by those who accept Maxwell's equations and reject the ether); and in precisely the way that biological phenomena and mental phenomena are accepted as primitive by vitalists and Cartesians. Semanticalism, like Cartesianism and vitalism, posits nonphysical primitives, and as a physicalist I believe that all three doctrines must be rejected.

There are two general sorts of strategy that can be taken in rejecting semanticalism, or Cartesianism, or vitalism. One strategy, illustrated two paragraphs back in discussing vitalism, is to try to explicate the terms of a biological theory in nonbiological terms. But there is another possible strategy, which is to argue that the biological terms are illegitimate. The second strategy seems reasonable to adopt in dealing with the following predicate of (reincarnationist) biology: ' $x$  has the same soul as  $y$ '. A physicalist would never try to find physical or chemical facts that underlie reincarnation; rather, he would reject reincarnation as a myth.

Since biological theory is as well developed as it is, we usually have a pretty good idea which biological terms require explication and which require elimination. When we turn to psychology and semantics, however, it is often not so obvious which strategy is the more promising. Thus in semantics, physicalists agree that all *legitimate* semantic terms must be explicable nonsemantically—they think in other words that there are no irreducibly semantic facts—but they disagree as to which semantic terms are legitimate. That disagreement has become fairly clear in recent years in the theory of meaning, with the work of Quine: the disagreement is between those physicalists who would look for a non-semantic basis for terms in the theory of meaning, and those who would follow Quine in simply throwing out those terms. Our concern, however, is not with the theory of meaning, but with the theory of reference, and here the disagreement has been less clear, since there haven't been many physicalists who openly advocate getting rid of terms like 'true' and 'denotes'. There were such physicalists in the early 1930s; part of the importance of Tarski's work was to persuade them that they were on the wrong track, to persuade them that we should explicate notions in the theory of reference nonsemantically rather than simply get rid of them.

The view that we should just stop using semantic terms (here and in the rest of this paper, I mean terms in the theory of reference, such as 'true'

and ‘denotes’ and ‘applies to’) draws its plausibility from the apparent difficulty of explicating these terms nonsemantically. People utter the sounds ‘Electrons have rest mass but photons don’t’, or ‘Schnee ist weiss und Gras ist grün’, and we apply the word ‘true’ to their utterances. We don’t want to say that it is a primitive and inexplicable fact about these utterances that they are true, a fact that cannot be explicated in non-semantic terms; this is as unattractive to a physicalist as supposing that it is a primitive and inexplicable fact about an organism at a certain time that it is in pain. But how could we ever explicate in nonsemantic terms the alleged fact that these utterances are true? *Part* of the explication of the truth of ‘Schnee ist weiss und Gras ist grün’, presumably, would be that snow is white and grass is green. But this would only be part of the explanation, for still missing is the connection between snow being white and grass being green on the one hand, and the German utterance being true on the other hand. It is this connection that seems so difficult to explicate in a way that would satisfy a physicalist, i.e., in a way that does not involve the use of semantic terms.

If, in face of these difficulties, we were ever to conclude that it was *impossible* to explicate the notions of truth and denotation in non-semantic terms, we would have either to give up these semantic terms or else to reject physicalism. It seems to me that that is essentially what Tarski is saying in the quotation at the end of the last section, and I have tried to make it plausible by sketching analogies to areas other than semantics. Tarski’s view, however, was that, for certain languages at least, semantic terms *are* explicable nonsemantically, and that truth definitions like T2 provide the required explication. It is understandable that as far as *philosophical* purposes go Tarski should think that T1 leaves something to be desired: after all, it merely explicates truth in terms of other semantic concepts; but what good does that do if those other concepts can’t be explicated nonsemantically? T2, then, has a strong *prima facie* advantage over T1. In the next section I will show that it is not a genuine advantage.

#### IV

The apparent advantage of T2 over T1, I have stressed, is that it appears to reduce truth to nonsemantic terms; and I *think* this is why Tarski

wanted to give a truth definition like T2 rather than like T1. This interpretation makes sense of Tarski's remark about physicalism, and it also explains why someone who was certainly not interested in "meaning analysis" as that is usually conceived would have wanted to give "definitions" of truth and would emphasize that, in these "definitions," "I will not make use of any semantical concept if I am not able previously to reduce it to other concepts." In any case, the problem of reducing truth is a very important problem, one which T1 and T2 provide a partial solution to, and one which T2 *might* be thought to provide a full solution to; and it is not at all clear what *other* interesting problems T2 could be thought to solve better than T1.

In Tarski's own exposition of his theory of truth, Tarski put very little stress on the problem of reduction or on any other problem with a clear philosophical or mathematical motivation; instead, he set up a formal criterion of adequacy for theories of truth without any serious discussion of whether or why this formal criterion is reasonable. Roughly, the criterion was this:<sup>14</sup>

(M) Any condition of the form

(2)  $(\forall e)[e \text{ is true} \equiv B(e)]$

should be accepted as an adequate definition of truth if and only if it is correct and ' $B(e)$ ' is a well-formed formula containing no semantic terms. (The quantifiers are to be taken as ranging over expressions of one particular language only.)

The "only if" part of condition M is not something I will contest. It rules out the possibility of T1 *by itself* being an adequate truth definition; and it is right to do so, if the task of a truth definition is to reduce truth to nonsemantic terms, for T1 provides only a *partial* reduction. (To complete the reduction we need to reduce primitive denotation to nonsemantic terms.) T2, on the other hand, meets condition M; so either T2 is superior to T1 as a reduction, or else condition M is too weak and the "if" part of it must be rejected. My own diagnosis is the latter, but the other possibility seems initially reasonable. After all, how could condition M be strengthened? We might try requiring that ' $B(e)$ ' be not only *extensionally* equivalent to ' $e$  is true', but *intensionally* equivalent to it; but this clearly won't do, for even if we grant that there is an intelligible

notion of intensional equivalence, our concern is not with analyzing the meaning of the word 'true' but with performing a reduction. A clear and useful standard of equivalence that is stronger than extensional equivalence but not so strong as to rule out acceptable reductions is unknown at the present time, so I know no way to improve on condition M. My view is that we have a rough but useful concept of reduction which we are unable to formulate precisely; but I must admit that the alternative view, that extensional equivalence is adequate, has an initial appeal.

A closer look, however, will reveal quite conclusively that extensional equivalence is not a sufficient standard of reduction. This can be seen by looking at the concept of valence. The valence of a chemical element is an integer that is associated with that element, which represents the sort of chemical combinations that the element will enter into. What I mean by the last phrase is that it is possible—roughly, at least—to characterize which elements will combine with which others, and in what proportions they will combine, merely in terms of their valences. Because of this fact, the concept of valence is a physically important concept, and so if physicalism is correct it ought to be possible to explicate this concept in physical terms—e.g., it ought to be possible to find structural properties of the atoms of each element that determine what the valence of that element will be. Early in the twentieth century (long after the notion of valence had proved its value in enabling chemists to predict what chemical combinations there would be) this reduction of the concept of valence to the physical properties of atoms was established; the notion of valence was thus shown to be a physicalistically acceptable notion.

Now, it would have been easy for a chemist, late in the last century, to have given a "valence definition" of the following form:

- (3)  $(\forall E)(\forall n)(E \text{ has valence } n \equiv E \text{ is potassium and } n \text{ is } +1, \text{ or } \dots, \text{ or } E \text{ is sulphur and } n \text{ is } -2)$

where in the blanks go a list of similar clauses, one for each element. But, though this is an extensionally correct definition of valence, it would not have been an acceptable reduction; and had it turned out that nothing else was possible—had all efforts to explain valence in terms of the structural properties of atoms proved futile—scientists would have eventually had to decide either (a) to give up valence theory, or else (b)

to replace the hypothesis of physicalism by another hypothesis (chemicalism?). It is part of scientific methodology to resist doing (b); and I also think it is part of scientific methodology to resist doing (a) as long as the notion of valence is serving the purposes for which it was designed (i.e., as long as it is proving useful in helping us characterize chemical compounds in terms of their valences). But the methodology is not to resist (a) and (b) by giving lists like (3); the methodology is to look for a real reduction. This is a methodology that has proved extremely fruitful in science, and I think we'd be crazy to give it up in linguistics. *And I think we are giving up this fruitful methodology, unless we realize that we need to add theories of primitive reference to T1 or T2 if we are to establish the notion of truth as a physicalistically acceptable notion.*

I certainly haven't yet given much argument for this last claim. I *have* argued that the standard of extensional equivalence doesn't guarantee an acceptable reduction; but T2 is obviously not trivial to the extent that (3) is. What *is* true, however, is roughly that T2 minus T1 is as trivial as (3) is. One way in which this last claim can be made more precise is by remembering that really we often apply the term 'valence' not only to elements, but also to configurations of elements (at least to stable configurations that are not compounds, i.e., to radicals). Thus, if we abstract from certain physical limitations on the size of possible configurations of elements (as, in linguistics, we usually abstract from the limitations that memory, etc., impose on the lengths of possible utterances), there is an infinite number of entities to which the term 'valence' is applied. But it is an important fact about valence that the valence of a configuration of elements is determined from the valences of the elements that make it up, and from the way they're put together. Because of this, we might try to give a recursive characterization of valence. First of all, we would try to characterize all the different *structures* that configurations of elements can have (much as we try to characterize all the different grammatical structures before we give a truth definition like T1 or T2). We would then try to find rules that would enable us to determine what the valence of a complicated configuration would be, given the valences of certain less complicated configurations that make it up and the way they're put together. If we had enough such rules, we could determine the valence of a given configuration given only its structure and the valences of the ele-



ments that make it up. And if we like, we can transform our recursive characterization of valence into an explicit characterization, getting

$$(V1) \quad (\forall c)(\forall n)(c \text{ has valence } n \equiv B(c, n))$$

The formula ' $B(c, n)$ ' here employed will still contain the term 'valence', but it will contain that term only as applied to elements, not as applied to configurations. Thus our "valence definition" V1 would characterize the valence of the complex *in terms of the valences of the simple*.

It would now be possible to eliminate the term 'valence' from ' $B(c, n)$ ', in either of two ways. One way would be to employ a genuine reduction of the notion of valence for elements to the structural properties of atoms. The other way would be to employ the pseudo-reduction (3). It is clear that we could use (3) to give a trivial reformulation V2 of V1, which would have precisely the "advantages" as a reduction that T2 has over T1. (V2, incidentally, would also have one of the disadvantages over V1 that T2 has over T1: V1 does not need to be overhauled when you discover or synthesize new elements, whereas V2 does.)

That is a sketch of one way that the remark I made two paragraphs back about "T2 minus T1" could be made more precise. But it is somewhat more fruitful to develop the point slightly differently: doing this will enable me to make clearer that there is unlikely to be *any* purpose that T2 serves better than T1 (not merely that T2 is no better at reduction).

To get this result I'll go back to my original use of the term 'valence', where it applies to elements only and not to configurations. And what I will do is compare (3) not to Tarski's theory of *truth*, but to Tarski's theory of *denotation* for names; the effect of this on his theory of truth will then be considered. Tarski states his theory of denotation for names in a footnote, as follows:

To say that the name  $x$  denotes a given object  $a$  is the same as to stipulate that the object  $a$  ... satisfies a sentential function of a particular type. In colloquial language it would be a function which consists of three parts in the following order: a variable, the word 'is' and the given name  $x$ . (CTFL 194)

This is actually only part of the theory, the part that defines denotation in terms of satisfaction; to see what the theory looks like when all semantic terms are eliminated, we must see how satisfaction is defined. The definition is given by the (A) and (B) clauses of T2, for, as I've remarked,

'satisfaction' is Tarski's name for what I've called "truth<sub>s</sub>". What Tarski's definition of satisfaction tells us is this: for any name  $N$ , an object  $a$  satisfies the sentential function  $\ulcorner x_1 \text{ is } N \urcorner$  if and only if  $a$  is France and  $N$  is 'France' or ... or  $a$  is Germany and  $N$  is 'Germany'. Combining this definition of satisfaction (for sentential functions of form  $\ulcorner x_1 \text{ is } N \urcorner$ ) with the earlier account of denotation in terms of satisfaction, we get:

- (DE) To say that the name  $N$  denotes a given object  $a$  is the same as to stipulate that either  $a$  is France and  $N$  is 'France', or ..., or  $a$  is Germany and  $N$  is 'Germany'.

This is Tarski's account of denotation for English proper names. For foreign proper names, the definition of denotation in terms of satisfaction needs no modification (except that the 'is' must be replaced by a name of a foreign word, say 'ist' for German). Combining this with the definition (again given by T2) of satisfaction for foreign sentential functions like  $\ulcorner x_1 \text{ ist } N \urcorner$ , we get:

- (DG) To say that the name  $N$  denotes a given object  $a$  is the same as to stipulate that either  $a$  is France and  $N$  is 'Frankreich', or ..., or  $a$  is Germany and  $N$  is 'Deutschland'.

DE and DG have not received much attention in commentaries on Tarski, but in fact they play a key role in his semantic theory; and it was no aberration on Tarski's part that he offered them as theories of denotation for English and German names, for *they satisfy criteria of adequacy exactly analogous to the criteria of adequacy that Tarski accepted for theories of truth*.<sup>15</sup> Nevertheless, it seems clear that DE and DG do not really reduce denotation to nonsemantic terms, any more than (3) reduces valence to nonchemical terms. What would a real explication of denotation in nonsemantic terms be like? The "classical" answer to this question (Russell's) is that a name like 'Cicero' is "analytically linked" to a certain description (such as 'the denouncer of Catiline'); so to explain how the name 'Cicero' denotes what it does you merely have to explain

- i. the process by which it is linked to the description (presumably you bring in facts about how it was learned by its user, or facts about what is going on in the user's brain at the time of the using) and
- ii. how the description refers to what it does

Because of (ii), of course, the project threatens circularity: the project is to explain how names refer in terms of how descriptions refer; but the natural way to explain how descriptions refer is in terms of how they're built up from their significant parts,<sup>16</sup> and how those significant parts refer (or apply, or are fulfilled), and those significant parts will usually include names. But Russell recognized this threat of circularity, and carefully avoided it: he assumed that the primitives of the language were to be partially ordered by a relation of "basicness," and that each name except a most basic ("logically proper") name was to be analytically linked to a formula containing only primitives more basic than it. The most basic primitives were to be linked to the world without the intervention of other words, by the relation of acquaintance.

This classical view of how names (and other primitives) latch onto their denotations is extremely implausible in many ways (e.g., it says you can refer only to things that are definable from "logically proper" primitives; it requires that there be certain statements, such as 'If Cicero existed then Cicero denounced Catiline', which are analytic in the sense that they are guaranteed by linguistic rules and are immune to revision by future discoveries). I conjecture that it is because of the difficulties with this classical theory, which was the only theory available at the time that Tarski wrote, that Tarski's pseudo-theories DE and DG seemed reasonable—they weren't exciting, but if you wanted something exciting you got logically proper names. The diagnosis that any attempt to explain the relation between words and the things they are about must inevitably lead to either a wildly implausible theory (like Russell's) or a trivial theory (like Tarski's) seems to be widely accepted still; but I think that the diagnosis has become less plausible in recent years through the development of *causal* theories of denotation by Saul Kripke<sup>17</sup> and others. According to such theories, the facts that 'Cicero' denotes Cicero and that 'muon' applies to muons are to be explained in terms of certain kinds of causal networks between Cicero (muons) and our uses of 'Cicero' ('muon'): causal connections both of a social sort (the passing of the word 'Cicero' down to us from the original users of the name, or the passing of the word 'muon' to laymen from physicists) and of other sorts (the evidential causal connections that gave the original users of the name "access" to Cicero and give physicists "access" to muons). I don't think that Kripke

or anyone else thinks that *purely* causal theories of primitive denotation can be developed (even for proper names of past physical objects and for natural-kind predicates); this however should not blind us to the fact that he has suggested a kind of factor involved in denotation that gives new hope to the idea of explaining the connection between language and the things it is about. It seems to me that the possibility of *some such* theory of denotation (to be deliberately very vague) is essential to the joint acceptability of physicalism and the semantic term 'denotes', and that denotation definitions like DE and DG merely obscure the need for this.

It might be objected that the purpose of DE and DG was not reduction; but what was their purpose? One answer might be that (DE) and (DG) enable us to eliminate the word 'denote' whenever it occurs. ("To explain is to show how to eliminate.") For instance,

- (4) No German name now in use denotes something that does not yet exist

would become

- (4') For any name *N* now in use, if *N* is 'Frankreich' then France already exists, and ..., and if *N* is 'Deutschland' then Germany already exists

provided that (DG) is a correct and complete list of the denotations of all those German proper names that have denotations. It seems reasonably clear that we could specify a detailed procedure for transforming sentences like (4) into materially equivalent sentences like (4'). A similar claim could be made for the "valence definition" (3). Such a valence definition makes it possible to eliminate the word 'valence' from a large class of sentences containing it, and in a uniform way. For instance,

- (5) For any elements *A* and *B*, if one atom of *A* combines with two of *B*, then the valence of *A* is  $-2$  times that of *B*.

is materially equivalent to

- (5') For any elements *A* and *B*, if one atom of *A* combines with two of *B*, then either *A* is sodium and *B* is sodium and  $+1 = -2 (+1)$ , or ..., or *A* is sulphur and *B* is sodium and  $-2 = -2 (+1)$ , or ...

provided that (3) is a correct and complete list of valences. So if anyone ever wants to eliminate the word ‘denote’ or the word ‘valence’ from a large class of English sentences by a uniform procedure, denotation definitions and valence definitions are just the thing he needs. There are, however, sentences from which these words are not eliminable by the sketched procedure. For instance, in semantics and possibly in chemistry there are problems with counterfactuals, e.g., ‘If ‘Germany’ had been used to denote France, then ...’. Moreover, there are special problems affecting the case of semantics, arising from the facts

- i. that the elimination procedure works only for languages in which nothing is denoted that cannot be denoted (without using semantic terms) in one’s own language,
- ii. that it works only for languages that contain no ambiguous names, and
- iii. that the denotation definitions provide no procedure for eliminating ‘denote’ from sentences where it is applied to more than one language; e.g., it gives no way of handling sentences like “‘Glub’ denotes different things in different languages.”

But, subject to these three qualifications (plus perhaps that involving counterfactuals), the elimination procedure for ‘denote’ is every bit as good as that for ‘valence’.

What value did Tarski attach to such transformations? Unfortunately he did not discuss the one about valences, but he did discuss the one that transforms “Smith used a proper name to denote Germany” into something logically equivalent to “Smith uttered ‘Deutschland’.” And it is clear that to this definition he attached great philosophical importance. After defining semantics as “the totality of considerations concerning those concepts which, roughly speaking, express certain connexions between the expressions of a language and the objects and states of affairs referred to by those expressions” (ESS 401), he says that with his definitions, “the problem of establishing semantics on a scientific basis is completely solved” (ESS 407). In other places his claims are almost as extravagant. For instance, the remark about physicalism that I quoted at the end of section II is intended to apply to denotation as well as to truth: if definitions of denotation like DE and DG could not be given, “it would ... be impossible to bring [semantics] into harmony with ... physi-

calism" (ESS 406); but because of these definitions, the compatibility of the semantic concept of denotation with physicalism is established. By similar standards of reduction, one might prove that witchcraft is compatible with physicalism, as long as witches cast only a finite number of spells: for then 'cast a spell' can be defined without use of any of the terms of witchcraft theory, merely by listing all the witch-and-victim pairs.

In other places Tarski makes quite different claims for the value of his denotation definitions. For example:

We desire semantic terms (referring to the object language) to be introduced into the meta-language only by definition. For, if this postulate is satisfied, the definition of truth, or of any other semantic concept [including denotation, which Tarski had already specifically mentioned to be definable], will fulfill what we intuitively expect from every definition; that is, it will explain the meaning of the term being defined in terms whose meaning appears to be completely clear and unequivocal.<sup>18</sup>

But it is no more plausible that DE "explains the meaning of" 'denote' as applied to English, or that DG "explains the meaning of" 'denote' as applied to German, than that (3) "explains the meaning of" 'valence'—considerably *less* so in fact, since for 'valence' there is no analogue to the conclusions that 'denote' means something different when applied to English than it means when applied to German. In fact, it seems pretty clear that denotation definitions like DE and DG have no philosophical interest whatever. But what conclusions can we draw from this about Tarski's *truth* definitions like T2? I think the conclusion to draw is that T2 *has no philosophical interest whatever that is not shared by* T1. How this follows I will now explain.

We have seen that Tarski advocated theories of denotation for names that had the form of mere lists: examples of his denotation definitions were DE and DG, and for language L his denotation definition would take the following form:

(D2)  $(\forall e)(\forall a)[e \text{ is a name that denotes } a \equiv (e \text{ is 'c}_1\text{' and } a \text{ is } \bar{c}_1) \text{ or } (e \text{ is 'c}_2\text{' and } a \text{ is } \bar{c}_2) \text{ or } \dots]$

where into the dots go analogous clauses for every name of L. Similarly, we can come up with definitions of application and fulfillment which are

acceptable according to Tarski's standards, and which also have the form of mere lists. The definition of application runs:

- (A2)  $(\forall e)(\forall a)[e \text{ is a predicate that applies to } a \equiv (e \text{ is 'p}_1\text{' and } \bar{p}_1(a)) \text{ or } (e \text{ is 'p}_2\text{' and } \bar{p}_2(a)) \text{ or } \dots].$

Similarly, we can formulate a list-like characterization F2 of fulfillment for the function symbols. Clearly neither A2 nor F2 is of any more theoretical interest than D2.

Tarski, I have stressed, accepted D2 as part of his semantic theory, and would also have accepted A2 and F2; and this fact is quite important, since D2, A2, and F2 together with T2 imply T1. In other words, T1 is simply a weaker version of Tarski's semantic theory; it is a logical consequence of Tarski's theory. Now, an interesting question is what you have to add to T1 to get the rest of Tarski's semantic theory. Suppose we can find a formula R that we can argue to be of no interest whatever, such that Tarski's semantic theory ( $T2 \wedge D2 \wedge A2 \wedge F2$ ) is logically equivalent to  $T1 \wedge R$ . It will then follow that the whole interest of Tarski's semantic theory lies in T1—the rest of his semantic theory results simply by adding to it the formula R, which (I have assumed) has no interest whatever. And if there is nothing of interest in the conjunction  $T2 \wedge D2 \wedge A2 \wedge F2$  beyond T1, certainly there can be nothing of interest in T2 alone beyond T1.

An example of such a formula R is  $D2 \wedge A2 \wedge F2$ : it is obvious that Tarski's semantic theory is logically equivalent to  $T1 \wedge D2 \wedge A2 \wedge F2$ . Because of this, *any interest in Tarski's semantic theory over T1 must be due to an interest in D2 or A2 or F2 (or to confusion): in this sense  $D2 \wedge A2 \wedge F2$  is "T2 minus T1"*. But I've already argued that D2, A2, and F2 have no theoretical interest whatever, and so that establishes that T2 has no theoretical interest whatever that is not shared by T1.

## V

Much of what I've said in this paper gains plausibility by being put in a wider perspective, and so I now want to say a little bit about why we want a notion of truth. The notion of truth serves a great many purposes, but I suspect that its original purpose—the purpose for which it was first

developed—was to aid us in utilizing the utterances of others in drawing conclusions about the world. To take an extremely simple example, suppose that a friend reports that he's just come back from Alabama and that there was a foot of snow on the ground there. Were it not for his report we would have considered it extremely unlikely that there was a foot of snow on the ground in Alabama—but the friend knows snow when he sees it and is not prone to telling us lies for no apparent reason, and so after brief deliberation we conclude that probably there *was* a foot of snow in Alabama. What we did here was first to use our evidence about the person and his situation to decide that he probably said something true when he made a certain utterance, and then to draw a conclusion from the truth of his utterance to the existence of snow in Alabama. In order to make such inferences, we have to have a pretty good grasp of (i) the circumstances under which what another says is likely to be true, and (ii) how to get from a belief in the truth of what he says to a belief about the extralinguistic world.

If this idea is right, then two features of truth that are intimately bound up with the purposes to which the notion of truth are put are (I) the role that the attempt to tell the truth and the success in doing so play in social institutions, and (II) the fact that normally one is in a position to assert of a sentence that it is true in just those cases where one is in a position to assert the sentence or a paraphrase of it. It would then be natural to expect that what is involved in communicating the meaning of the word 'true' to a child or to a philosopher is getting across to him the sorts of facts listed under (I) and (II); for those are the facts that it is essential for him to have an awareness of if he is to put the notion of truth to its primary use (child) or if he is to get a clear grasp of what its primary use is (philosopher).

I think that this natural expectation is correct, and that it gives more insight than was given in sections II and IV into why it is that neither T1 nor T2 can reasonably be said to explain the meaning of the term 'true'—even when a theory of primitive reference is added to them. First consider (I). The need of understanding the sort of thing alluded to in (I), if we are to grasp the notion of truth, has been presented quite forcefully in Michael Dummett's article "Truth,"<sup>19</sup> in his analogy between speaking the truth and winning at a game. It is obvious that T1 and T2 don't



explain anything like this (and in fact Dummett's fourth paragraph, on Frege-style truth definitions, can be carried over directly to T1 and T2).

The matter might perhaps be expressed in terms of assertibility conditions that one learns in learning to use the word 'true': part of what we learn, in learning to use this word, is that in cases like that involving the friend from Alabama there is some *prima facie* weight to be attached to the claim that the other person is saying something true. But there are also *other* assertibility conditions that one learns in learning the word 'true', assertibility conditions which have received considerable attention in the philosophical literature on truth. To begin with, let's note one obvious fact about how the word 'true' is standardly learned: we learn how to apply it to utterances of our own language first, and when we later learn to apply it to other languages it is by conceiving the utterances of another language more or less on the model of utterances of our own language. The obvious model of the first stage of this process is that we learn to accept all instances of the schema

(T) *X* is true if and only if *p*

where '*X*' is replaced by a quotation-mark name of an English sentence *S* and '*p*' is replaced by *S*. This must be complicated to deal with ambiguous and truth-value-less sentences, but let's ignore them. Also let's ignore the fact that certain pathological instances of (T)—the Epimenides-type paradoxical sentences—are logically refutable. Then there is a sense in which the instances of (T) that we've learned to assert determine a unique extension for the predicate 'true' as applied to sentences of our own language.<sup>20</sup> Our views about what English sentences belong to this unique extension may be altered, but as long as we stick to the instances of (T) they cannot consistently be altered without also altering our beliefs in what those sentences express. This fact is extremely important to the functions that the word 'true' serves (as the Alabama example illustrates).

In stressing the assertibility conditions for simple sentences containing the word 'true', I have followed Quine (*ibid.* 138); for, like him, I believe that such assertibility conditions are enough to make the term 'true' reasonably clear. But now it might be asked, "Then why do we need causal (etc.) theories of reference? The words 'true' and 'denotes' are made perfectly clear by schemas like (T). To ask for more than these schemas—

to ask for causal theories of reference to nail language to reality—is to fail to recognize that we are at sea on Neurath's boat: we have to work *within* our conceptual scheme, we can't glue it to reality from the outside."

I suspect that this would be Quine's diagnosis—it is strongly suggested by §6 of *Word and Object*, especially when that is taken in conjunction with some of Quine's remarks about the inscrutibility of reference and truth value, the underdetermination of theories, and the relativity of ontology. It seems to me, however, that the diagnosis is quite wrong. In looking for a theory of truth and a theory of primitive reference we *are* trying to explain the connection between language and (extralinguistic) reality, but we are *not* trying to step outside of our theories of the world in order to do so. Our accounts of primitive reference and of truth are not to be thought of as something that could be given by philosophical reflection prior to scientific information—on the contrary, it seems likely that such things as psychological models of human beings and investigations of neurophysiology will be very relevant to discovering the mechanisms involved in reference. *The reason why accounts of truth and primitive reference are needed is not to tack our conceptual scheme onto reality from the outside; the reason, rather, is that without such accounts our conceptual scheme breaks down from the inside.* On our theory of the world it would be extremely surprising if there were some non-physical connection between words and things. Thus if we could argue from our theory of the world that the notion of an utterer's saying something true, or referring to a particular thing, cannot be made sense of in physicalist terms (say, by arguing that any semantic notion that makes physicalistic sense *can* be explicated in Skinnerian terms, and that the notions of truth and reference *can't* be explicated in Skinnerian terms), then to the extent that such an argument is convincing we ought to be led to conclude that, if we are to remain physicalists, the notions of truth and reference must be abandoned. No amount of pointing out the clarity of these terms helps enable us to escape this conclusion: 'valence' and 'gene' were perfectly clear long before anyone succeeded in reducing them, but it was their reducibility and not their clarity before reduction that showed them to be compatible with physicalism.

The clarity of 'valence' and 'gene' before reduction—and even more, their *utility* before reduction—did provide physicalists with substantial

reason to think that a reduction of these terms was possible, and, as I remarked earlier, a great deal of fruitful work in physical chemistry and chemical genetics was motivated by the fact. Similarly, insofar as semantic notions like 'true' are useful, we have every reason to suspect that they will be reducible to nonsemantic terms, and it is likely that progress in linguistic theory will come by looking for such reductions. (In fact, the fruitfulness of Tarski's work in aiding us to understand language is already some sign of this, even though it represents only a partial reduction.) Of course, this sort of argument for the prospects of reducing semantic notions is only as powerful as our arguments for the utility of semantic terms; and it is clear that the question of the utility of the term 'true'—the purposes it serves, and the extent to which those purposes could be served by less pretentious notions such as warranted assertibility—needs much closer investigation.

All these remarks require one important qualification. The notion of valence, it must be admitted, is *not* reducible to nonchemical terms on the *strictest* standards of reduction, but is only *approximately* reducible; yet, in spite of this, we don't want to get rid of the notion, since it is still extremely useful in those contexts where its approximate character isn't too likely to get in the way and where if we did not approximate we'd get into quantum-mechanical problems far too complex for anyone to solve. (Moreover, considerations about the purposes of the notion of valence were sufficient to show that the notion of valence would only be approximately reducible: for the utility of the notion of valence is that it aids us in approximately characterizing which elements will combine with which and in what proportions; yet it is obvious that no *precise* such characterization is possible.)

Similarly, it may well be that a detailed investigation into the purposes of the notion of truth might show that these purposes require only an approximate reduction of the notion of truth. Still, to require an approximate reduction is to require quite a bit; after all, 'is a reincarnation of' isn't even approximately reducible to respectable biology, and 'electromagnetic field' is not approximately reducible to mechanics. Obviously the notion of approximate reduction needs to be made more precise (as in fact does the notion of strict, or nonapproximate, reduction); but even

without making it so, I think we can see that T2 is no more of an approximate reduction than is V2, since  $D2 \wedge A2 \wedge F2$  is no more of an approximate reduction than is (3). In other words, the main point of the paper survives when we replace the ideal of strict reduction by the ideal of approximate reduction.

It should be kept carefully in mind that the Quinean view that all we need do is clarify the term 'true', in the sense that this term is clarified by schema T (or by schema T plus a theory of translation to handle foreign languages; or by schema T plus the sort of thing alluded to in connection with Dummett), is *not* Tarski's view. Tarski's view is that we have to provide a truth characterization like T2 (which, when we choose as our object language L a "nice" fragment of our own language, can be shown correct merely by assuming that all instances of schema T are valid—cf. fn 14, p. 361); and such a truth characterization does much more than schema T does. It does not do everything that Tarski ever claimed for it, for Tarski attached much too much importance to the pseudo-theories D2, A2, and F2; but even when we "subtract" such trivialities from his truth characterization T2, we still get the very interesting and important truth characterization T1. T1, I believe, adequately represents Tarski's real contribution to the theory of truth, and in doing this it has a number of positive advantages over T2 (in addition to the important negative advantage I've been stressing, of preventing extravagant claims based on the fact that T2 contains no semantic terms). First of all, T1, unlike T2, is applicable to languages that contain ambiguities and languages that contain terms not adequately translatable into English. Second, T1, unlike T2, can be used in diachronic linguistics: it doesn't need overhauling as you add new words to the language, provided those new words belong to the same semantic category as words already in the language. Third, I think that the reason why Tarski's theory of truth T2 has seemed so uninteresting to so many people is that it contains the vacuous semantic theories D2, A2, and F2 for the primitives of the language. By expressing the really important features of Tarski's results on truth, and leaving out the inessential and uninteresting "theories" of the semantics of the primitives, T1 should make the philosophical importance of Tarski's work more universally recognized.

## Notes

This paper grew out of a talk I gave at Princeton in the fall of 1970, where I defended T1 over T2. Donald Davidson and Gilbert Harman—and later, in private conversation, John Wallace—all came to the defense of T2, and their remarks have all been of help to me in writing the paper. I have also benefited from advice given by Michael Devitt, Paul Benacerraf, and especially David Hills.

1. *Logic of Scientific Discovery* (New York: Basic Books, 1968), p. 274.
2. Cf. Carnap's "Autobiography," in P. A. Schilpp, ed., *The Philosophy of Rudolf Carnap* (Lasalle, Ill.: Open Court, 1963), p. 61.
3. It is sometimes claimed that Tarski was interested in languages considered in abstraction from all speakers and writers of the language; that the languages he was dealing with are abstract entities to be specified by giving their rules. This seems incorrect: Tarski was interested in giving the semantics of languages that mathematicians had been writing for years; and only as a result of Tarski's work was it then possible for philosophers like Carnap to propose that the clauses of a Tarski-type truth definition for such languages be called rules of the languages and be used in defining the languages as abstract entities.
4. "The Concept of Truth in Formalized Languages" (CTFL), in *Logic, Semantics, and Metamathematics (LSM)* (New York: Oxford, 1956), p. 166.
5. Actually in model theory we are interested in allowing a slightly unusual semantics for the quantifiers: we are willing to allow that the quantifier not range over everything. We could build this generalization into our truth definition, by stipulating that in addition to the denotations of the nonlogical symbols we specify a universe  $U$ , and then reformulating clause (B)4 by requiring that the  $k$ th member of  $s^*$  belong to  $U$ . If we did this, then it would be the range of the quantifiers as well as the denotations of the nonlogical primitives that we would have explained truth in terms of.
6. An *incomplete* sentence token is a sentence token which [like the occurrence of ' $2 + 2 = 4$ ' inside ' $\sim (2 + 2 = 4)$ '] is part of a larger sentence token.
7. "Truth and Meaning," *Synthese*, xvii, 3 (September, 1967): 304–323, pp. 314/5.
8. Note that the claims I've been making are intended to apply only to cases where different tokens have different semantic features; they are not intended to apply to cases of indeterminacy, i.e., to cases where a particular name token or predicate token has no determinate denotation or extension. To deal with indeterminacy requires more complex devices than I employ in this paper.
9. The notion of a semantic category is Tarski's: cf. CTFL, p. 215.
10. To do so in the obvious way requires that we introduce semantic categories of negation symbol, conjunction symbol, and universal-quantification symbol; though by utilizing some ideas of Frege it could be shown that there is really no need of a separate semantic category for each logical operator. The use of semantic

categories in the generalized truth characterization raises important problems which I have had to suppress for lack of space in this paper.

11. For simplicity, I have assumed that L itself contains no semantic terms.

12. The Establishment of Scientific Semantics" (ESS) in *LSM*, p. 406.

13. This, of course, is very vague, but most attempts to explicate the doctrine of physicalism more precisely result in doctrines that are very hard to take seriously [e.g., the doctrine that for every acceptable predicate ' $P(x)$ ' there is a formula ' $B(x)$ ' containing only terminology from physics, such that ' $\forall x(P(x) \equiv B(x))$ ' is true]. Physicalism should be understood as the doctrine (however precisely it is to be characterized) that guides science in the way I describe.

14. Tarski actually gives a different formulation, the famous Convention T, evidently because he does not think that the word 'correct' ought to be employed in stating a criterion of adequacy. First of all Tarski writes

... we shall accept as valid every sentence of the form

[T] The sentence  $x$  is true if and only  $p$

where ' $p$ ' is to be replaced by any sentence of the language under investigation and ' $x$ ' by any individual name of that sentence provided this name occurs in the metalanguage (ESS 404).

Is Tarski's policy of accepting these sentences as "valid" (i.e., true) legitimate? It seems to me that it is, in a certain special case. The special case is where

I. The object language is a proper part of the metalanguage (here, English).

II. The object language contains no paradoxical or ambiguous or truth-value-less sentences.

In this special case—and it was the case that Tarski was primarily concerned with—I think it will be generally agreed that all instances of Schema T hold. From this, together with the fact that only grammatical sentences are true, we can argue that, if a necessary and sufficient condition of form (2) has the following consequences:

(a) Every instance of Schema T

(b) The sentence ' $(\forall x)(x \text{ is true} \supset S(x))$ ', where ' $S(x)$ ' formulates (correct) conditions for an utterance of L to be a sentence

then that necessary and sufficient condition is correct. Let's say that a "truth definition" for L (a necessary and sufficient condition of truth in L) *satisfies Convention T* if it has all the consequences listed under (a) and (b). Then, restating: when L is a language for which I and II hold, then any truth definition satisfying Convention T is correct; and since only quite uncontroversial assumptions about truth are used in getting this result, anyone will admit to the correctness of a truth characterization satisfying Convention T. If we use the term 'formally correct definition' for a sentence of form (2) in which ' $B(e)$ ' contains no semantic terms, this means that a formally correct definition that satisfies Convention T is bound to satisfy Condition M (when the language L satisfies I and II). As far as I

can see, this is the only motivation for Convention T; if so, we can discredit Convention T by discrediting Convention M.

Tarski sometimes states a more general form of Convention T, which applies to languages that do not meet restriction I: it is what results when one allows as instances of Schema T the results of replacing '*p*' by a *correct translation* of the sentence that the name substituted for '*x*' denotes (in some sense of 'correct translation' in which correctness requires preservation of truth value). But then the advantage of the ungeneralized form of Convention T (viz., that anything satisfying it wears its correctness on its face, or more accurately, on the faces of its logical consequences) is lost.

15. A sentence of the form ' $(\forall N)(\forall x)[N \text{ denotes } x \equiv B(N, x)]$ ' satisfies *convention* D if it has as consequences every instance of the schema '*y* denotes *z*', in which '*y*' is to be replaced by a quotation-mark name for a name *N*, and '*z*' is to be replaced by (an adequate translation of *N* into English, i.e.) a singular term of English that contains no semantic terms and that denotes the same thing that *N* denotes. Clearly DE and DG are not only extensionally correct, they also satisfy Convention D. Presumably philosophers who are especially impressed with Convention T will be equally impressed with this fact, but they owe us a reason why satisfying Convention D is of any interest.

16. For example, by extending our definition of denotation<sub>s</sub> to descriptions by:

$\ulcorner ix_k(e) \urcorner$  denotes<sub>s</sub> *a* if and only if [for each sequence *s*\* which differs from *s* at the *k*th place at most, *e* is true<sub>s</sub>\* if and only if the *k*th member of *s*\* is *a*].

and then defining denotation in terms of denotation<sub>s</sub> by stipulating that a closed term denotes an object if and only if it denotes<sub>s</sub> that object for some (or all) *s*.

17. Some of Kripke's work on names will be published shortly in Davidson and Harman, eds., *Semantics of Natural Language* (Dordrecht: Reidel, 1971). What I've said about Russell's view is influenced by some of Kripke's lectures on which his paper there is based.

18. "The Semantic Conception of Truth and the Foundations of Semantics," *Philosophy and Phenomenological Research*, iv, 3 (March 1944): 341–375, p. 351.

19. *Proceedings of the Aristotelian Society*, LIX (1958/9): 141–162. (Reprinted as chap. 10—Ed.)

20. Cf. W. V. Quine, *From a Logical Point of View* (New York: Harper & Row, 1961), p. 136.

---

## What Is a Theory of Truth?

Scott Soames

### I

Alfred Tarski's theory of truth and its successors enjoy a perplexing double status. On the one hand, they are mathematical theories characterized by a rich class of mathematical results. On the other hand, they are commonly believed by philosophers to provide analyses of the nature of truth and, hence, to be philosophically significant. With this broader significance comes a kind of controversy not normally associated with mathematical theorems. No one disputes the correctness of Tarski's formal results. In that sense, there is no doubt that his theory is true. However, there is considerable doubt about whether, or in what sense, it is a theory of truth.

One main reason for this uncertainty is the difficulty of determining what a theory of truth ought to be. Generally, theories of truth have tried to do one or the other of three main things:

- i. to give the meaning of natural-language truth predicates;
- ii. to replace such predicates with substitutes, often formally defined, designed to further some reductionist program; or
- iii. to use some antecedently understood notion of truth for broader philosophical purposes, such as explicating the notion of meaning or defending one or another metaphysical view.

In order to do the first of these things, a theory must analyze the content of paradigmatic examples in which what is said to be true is a proposition, rather than a sentence or utterance.



- (1) a. The proposition that the earth moves is true.
- b. Church's theorem is true.
- c. Everything he said is true.

There are theories that try, in my opinion unsuccessfully, to do just this.<sup>1</sup> Tarski's theory, which restricts itself to cases in which truth is predicated of sentences of certain formal languages, is not one of them. Thus, Tarski cannot be seen as even attempting to give the meaning of natural-language truth predicates.

Nor can he be seen as attempting to use the notion of truth for broad philosophical purposes. In order to do that, one must regard truth as essentially unproblematic and philosophically productive. For Tarski, truth itself is what has to be legitimated. Once it is, it turns out to be useful for certain primarily technical purposes, but useless for ambitious philosophical programs. For example, Tarski recognized that his notion of truth could not be used to give the meanings of logical constants (or, I suspect, anything else).<sup>2</sup> He also thought of it as epistemologically and metaphysically neutral. Thus, in "The Semantic Conception of Truth" he says:

... the semantic definition of truth implies nothing regarding the conditions under which a sentence like ...

Snow is white

can be asserted. It implies only that, whenever we assert or reject this sentence, we must be ready to assert or reject the correlated sentence ...

The sentence "Snow is white" is true.

Thus, we may accept the semantical conception of truth without giving up any epistemological attitude we may ever have had; we may remain naive realists, critical realists or idealists, empiricists or metaphysicians—whatever we were before. The semantic conception is completely neutral toward all these issues. (P. 356 herein)

It is helpful, in understanding this remark, to focus on something that the truth predicate is good for—namely, what W. V. Quine has called "semantic ascent."<sup>3</sup> The simplest example of this is provided by Tarski:

- (2) a. Snow is white.
- b. The sentence 'Snow is white' is true.

Any speaker of English knows that these sentences are at least materially equivalent. Because of this, they can often be used to convey essentially the same information. To choose (2b) is to use a semantic statement to convey information that could have been conveyed nonmetalinguistically. To do this is to engage in semantic ascent.

The importance of semantic ascent is illustrated by cases like (3), in which we want to generalize.

- (3) a. Snow is white  $\rightarrow$  (Grass is blue  $\rightarrow$  Snow is white)
- b. The earth moves  $\rightarrow$  (The sun is cold  $\rightarrow$  The earth moves)
- $\vdots$

Each of these examples is something one could feel safe in asserting. However, if one wanted to get the effect of asserting all of them, one would have to quantify, replacing sentences with variables. In English such quantification is most naturally, though not inevitably, construed as first-order and objectual. Thus, if the variables are taken to range over sentences we need a metalinguistic truth predicate. Semantic ascent gives us

- (4) For all sentences  $p, q$  ( $p$  is true  $\rightarrow$  ( $q$  is true  $\rightarrow p$  is true))<sup>4</sup>

That which is conveyed by (4) is closely related to that which is conveyed by (3). However, here the truth predicate is especially handy, since we don't have the alternative of asserting each member of (3).

Truth predicates can be used in the same way in more obviously philosophical cases. For example, consider:

- (5) There is a duplicate of our sun in some remote region of space, but we will never find (sufficient) evidence that there is.

Someone who asserted this would, by contemporary standards, be counted a metaphysical realist—i.e. as being someone who thinks that what there is doesn't depend in any way on what we may rationally believe. Of course, one can be a realist without believing (5). One may think that what there is doesn't depend on us, while believing that there is no duplicate of our sun, or being uncertain whether there is, or while holding that evidence will someday be found to settle the matter.

What, then, distinguishes realism from anti-realism? One is tempted to answer that it is the belief that

- (6) Either there is a duplicate of our sun in some remote region of space, but we will never find (sufficient) evidence that there is; or there is no duplicate of our sun in any remote region of space, but we will never find (sufficient) evidence that there isn't; or there is intelligent life elsewhere in the universe, but we will never find (sufficient) evidence that there is; or  
 ⋮

But this is awkward. We ought to be able to state the realist's position without having to gesture toward an infinite list. Semantic ascent provides a convenient way of doing this. With the help of a truth predicate and quantification over sentences, we can characterize the realist as believing, and the anti-realist as denying:

- (7) There is at least one sentence *s* such that *s* is true (in English), but we will never find (sufficient) evidence supporting *s*.

The relationship between (6) and (7) is like that between (2a) and (2b) and between (3) and (4). In each case, the semantic sentence may not say exactly the same thing as its nonsemantic counterpart; but if knowledge of English is assumed, the two can be used to convey essentially the same information.<sup>5</sup>

The utility of the truth predicate in stating this dispute has led some to believe that the dispute is about truth and, hence, that truth is a deeply metaphysical notion. However, there is no reason to suppose this. The realist and anti-realist may agree about truth; they may even accept something like Tarski's definition. Where they differ is in their conceptions of reality. Since statements about truth mirror direct statements about nonlinguistic reality, semantic ascent makes the truth predicate a convenient vehicle for expressing competing metaphysical views. But a convenient vehicle is all it is. As Tarski puts it, the notion of truth is completely neutral toward all these issues.

The upshot of this is that Tarski's definition of truth is neither an attempt to analyze the meaning of natural-language truth predicates nor an attempt to use the notion of truth for broad philosophical purposes. Rather, Tarski's goal was to replace natural-language truth predicates with certain restricted, but formally defined substitutes. He thought such replacements were needed both to remove the doubts of certain scientific

cally minded truth skeptics and to eliminate what he took to be the incoherence in our ordinary notion brought out by the liar paradox.

For Tarski, these two motivations were connected, since the paradoxes constituted one source of skepticism about truth.<sup>6</sup> However, the truth skeptics of his day also had other, more broadly philosophical grounds for their doubts. These included the frequent use of truth in metaphysical discussions, the tendency to confuse truth with epistemological notions like certainty and confirmation, and the inability to see how acceptance of a truth predicate could be squared with the doctrine of physicalism and the unity of science.<sup>7</sup> Although Tarski's work was historically effective in alleviating each of these worries, the only one discussed by Tarski was the final one, involving physicalism.<sup>8</sup>

Tarski's version of physicalism was a moderate one, allowing both physical and mathematical elements, without requiring the latter to be reduced to the former. Roughly, this "moderate physicalism" asserts that

- i. all facts are physical or mathematical facts;
- ii. all scientific (or descriptive) claims are reducible to claims about the physical and mathematical characteristics of things; and
- iii. all scientific (or descriptive) concepts are definable in terms of physical and mathematical concepts.<sup>9</sup>

Tarski took this doctrine to require that truth be eliminable via an explicit, physicalistic definition. Anything else—for example, taking truth to be a primitive whose extension is fixed by a set of axioms—was deemed to be undesirable.

It is worth pointing out that this emphasis on definition is primarily philosophical rather than technical. What is at issue is not the technical results achievable, but the philosophical significance of those results.<sup>10</sup> It is possible to view a Tarski truth characterization for a language *L* as simply specifying the extension of 'true' for *L*, explaining how the truth value of a sentence depends on the semantic properties of its parts, and providing the basis for accounts of logical truth and logical consequence. Even if the truth characterization is put in the form of what is technically an explicit definition, it doesn't have to be viewed as an explication of truth in any interesting philosophical sense. If one's philosophical views differ from Tarski's, one can accept his formal results while taking truth to be primitive.

I will not comment directly on this way of viewing Tarski, but will instead concentrate on his own view of his work. I do this not out of any commitment to physicalism, but rather out of a sense that his deflationist attitude toward truth is interesting and worth defending. Tarski's attitude needs defense because his definition of truth fails to satisfy certain initially plausible demands one might place on an explication of truth. His attitude is defensible because these demands turn out to be dubious or illegitimate. The importance of this defense extends beyond Tarski to the general question of what ought, and what ought not, to be expected from a theory of truth.

## II

Tarski's basic idea is that for certain languages  $L$ , adequate for natural science, one can define a truth predicate using only notions already expressible in  $L$ , plus certain syntactic and set-theoretic apparatus. Thus, if  $L$  is physicalistically pure and if syntax and set theory are unproblematic, then defining a metalanguage truth predicate can't introduce any difficulties.

Following Hartry Field, we can think of such a definition as divided into two parts. The first part is concerned with what Field calls "primitive denotation"; here one defines what it is for a name to refer to an object and for a predicate to apply to one or more objects. The second part of the definition defines truth in terms of primitive denotation. The end result is a metalanguage sentence:

(8) For all sentences  $s$  of  $L$ ,  $s$  is true iff  $T(s)$ .

in which  $T(s)$  is a formula with only ' $s$ ' free, made up entirely of logical, set-theoretic, and syntactic apparatus, plus translations of the primitives of  $L$ . If these translations are (extensionally) correct, then  $T(s)$  will be coextensive with 'true' over  $L$ .

Tarski's technique can be illustrated using a particularly simple example. Let  $L$  be a language whose only logical constants are ' $\vee$ ' and ' $-$ ', and whose nonlogical constants consist of finitely many names and one-place predicates. (R) and (A) define reference and application for  $L$ ; (T) uses these notions to define truth:

- (R) For all names  $n$  of  $L$  and objects  $o$ ,  $n$  refers in  $L$  to  $o$  iff  $n = 'a'$  and  $o = \text{Arizona}$ , or  $n = 'b'$  and  $o = \text{Boston}$ , ... (and so on for each of the names of  $L$ ).
- (A) For all one-place predicates  $P$  of  $L$  and objects  $o$ ,  $P$  applies in  $L$  to  $o$  iff  $P = 'C'$  and  $o$  is a city, or  $P = 'S'$  and  $o$  is a state, ... (and so on for each one-place predicate of  $L$ ).
- (T) For all sentences  $S$  of  $L$ ,  $S$  is true in  $L$  iff  $S \in$  the set  $K$  such that for all  $x$ ,  $x \in K$  iff
  - i.  $x = \ulcorner Pn \urcorner$  for some predicate  $P$  and name  $n$  of  $L$ , and there is an object  $o$  such that  $n$  refers in  $L$  to  $o$  and  $P$  applies in  $L$  to  $o$ ; or
  - ii.  $x = \ulcorner (A \vee B) \urcorner$  for some formulas  $A$  and  $B$  of  $L$ , and  $A \in K$  or  $B \in K$ ; or
  - iii.  $x = \ulcorner \neg A \urcorner$  for some formula  $A$  of  $L$  and  $A \notin K$ .

Let (T') be just like (T) except for containing the right-hand sides of (R) and (A) where (T) contains  $n$  refers in  $L$  to  $o$  and  $P$  applies in  $L$  to  $o$ , respectively. (T') is then an explicit Tarskian truth definition for  $L$ , with 'T(s)' in (8) representing the right-hand side of (T').<sup>11</sup>

Although truth definitions for richer languages are technically more interesting, their philosophical status as putative physicalistic reductions of truth is essentially the same as that of the simple definition just given. On the basis of such definitions, Tarski concluded that he had shown truth, reference, and application to be physicalistically acceptable terms.

In a well-known critique of Tarski, Hartry Field (op. cit.) argues that this conclusion is unjustified. The problem, according to Field, is that the proposed replacements for the notions of primitive denotation are not physicalistically acceptable reductions of our pretheoretic notions of reference and application. Because Field takes Tarski to have reduced truth to primitive denotation (P. 368 herein), he concludes that Tarski has not legitimated the notion of truth for physicalists.

Field does not, of course, dispute the fact that Tarski's definitions are extensionally correct. He maintains, however, that extensional correctness is not enough. In addition, any genuine reduction must show semantic facts about expressions to be supervenient on physical facts about their users and the environments in which they are used. Tarski's definitions don't do this.

This can be seen by considering a simple example. Suppose that '*Cb*' is a sentence of *L* and that the relevant semantic facts about it are given in (9):

- (9) a. '*b*' refers (in *L*) to Boston.  
 b. '*C*' applies (in *L*) to cities (and only cities).  
 c. '*Cb*' is true (in *L*) iff Boston is a city.

If Tarski's definitions really specify the physicalistic content of semantic notions, then, in each case, we ought to be able to substitute the physicalistic definiens for the semantic definiendum without changing the physical fact thereby specified. Performing this substitution and simplifying results, we obtain

- (10) a. '*b*' = '*b*' and Boston = Boston.  
 b. For all objects *o*, '*C*' = '*C*' and *o* is a city, iff *o* is a city.  
 c. '*Cb*' = '*Cb*' and there is an object *o* such that *o* = Boston and *o* is a city, iff Boston is a city.

But there is a problem in identifying these facts with those in (9). As Field points out, it is natural to suppose that the expressions of a language have semantic properties only in virtue of the ways they are used by speakers. Thus, he holds that the facts given in (9) wouldn't have obtained if speakers' linguistic behavior had been different.<sup>12</sup> Since the facts in (10) are not speaker-dependent in this way, Field concludes that they are not semantic facts and that Tarski's attempted reduction fails. Tarski's truth predicate is both physicalistic and coextensive with 'true in *L*'; but it is not, according to Field, a physicalistic conception of truth.

On Field's view, Tarski's truth characterization inherits its inadequacy as a reduction from the pseudo-reductions that constitute its base clauses. Thus, Field's strategy for solving the problem is to provide genuine reductions for the notions of primitive denotation, on something like the model of the causal theory of reference. The picture that emerges from his discussion is one in which an adequate definition of truth is a two-stage affair. Stage 1 is Tarski's reduction of truth to primitive denotation. Stage 2 is the imagined causal theory-like reduction of the notions of a name referring to, and a predicate applying to, an object in a language.<sup>13</sup> If the physical facts that determine denotation in one language do so in all, then these relations will hold between expressions and objects, for

variable ' $L$ '. When logical vocabulary and syntax is kept fixed, the result is a notion of truth that is not language-specific, but is itself defined for variable ' $L$ '.

Although the resulting picture appears rosy, there are several problems with it. One concerns reference to abstract objects, for which a causal account seems problematic. Another involves Quinean worries about ontological relativity and referential indeterminacy. These, of course, are obstacles to a physicalistic reduction of primitive denotation. However, there are other difficulties which become clear when one notices that Field has understated his objection to Tarski. If the alleged dependence of semantic facts on facts about speakers shows that Tarski has not reduced primitive denotation to physical facts, then the very same point shows that he has not reduced truth to primitive denotation.

This can be seen by considering a pair of elementary examples. Imagine two languages,  $L_1$  and  $L_2$ , which are identical except that in  $L_1$  the predicate ' $R$ ' applies to round things, whereas in  $L_2$  it applies to red things. Owing to this difference, certain sentences will have different truth conditions in the two languages.

- (11) a. ' $Re$ ' is true in  $L_1$  iff the earth is round.
- b. ' $Re$ ' is true in  $L_2$  iff the earth is red.

Under Tarski's original definition, this difference will be traceable to the base clauses of the respective truth definitions, where the applications of predicates are simply listed.

Field's objection to this is that although Tarski's definitions correctly *report* that ' $R$ ' applies to different things in the two languages, they don't *explain* how this difference arises from the way in which speakers of the two languages use the predicate. What Field fails to point out is that exactly the same objection can be brought against Tarski's treatment of logical vocabulary and syntax in the recursive part of his definition.

This time let  $L_1$  and  $L_2$  be identical except for their treatment ' $\vee$ '.

- (12) a. A formula  $\ulcorner(A \vee B)\urcorner$  is true in  $L_1$  (with respect to a sequence  $s$ ) iff  $A$  is true in  $L_1$  (with respect to  $s$ ) or  $B$  is true in  $L_1$  (with respect to  $s$ ).
- b. A formula  $\ulcorner(A \vee B)\urcorner$  is true in  $L_2$  (with respect to a sequence  $s$ ) iff  $A$  is true in  $L_2$  (with respect to  $s$ ) and  $B$  is true in  $L_2$  (with respect to  $s$ ).



Owing to this difference, sentences containing 'v' will have different truth conditions in the two languages. In order to satisfy Field's requirements on reduction, it is not enough for a truth characterization to report such differences. Rather, such differences must be explained in terms of the manner in which speakers of the two languages treat 'v'.<sup>14</sup> Since Tarski's truth definitions don't say anything about this, their recursive clauses should be just as objectionable to the physicalist as the base clauses.

This means that Field's strategy of achieving a genuine reduction of truth by supplementing Tarski with nontrivial definitions of primitive denotation cannot succeed. The reason it can't is that, given Field's strictures on reduction, Tarski has not reduced truth (for standard first-order languages) to primitive denotation. At best he has reduced it to the class of semantic primitives listed in (13):<sup>15</sup>

- (13) the notion of a name referring to an object
- the notion of a predicate applying to objects
- the notion of a formula being the application of an  $n$ -place predicate  $P$  to an  $n$ -tuple of terms  $t_1 \dots t_n$
- the notion of a formula  $A$  being a negation of a formula  $B$
- the notion of a formula  $A$  being a disjunction of formulas  $B$  and  $C$
- the notion of a formula  $A$  being an existential generalization of a formula  $B$  with respect to a variable  $u$  and a domain  $D$  of objects

This way of looking at things requires a restatement of every clause in Tarski's truth definition. For example, the recursive clause for negation, which had been given by (14a), is now given by (14b).

- (14) a. If  $A = \neg B$ , then  $A$  is true in  $L$  (with respect to a sequence  $s$ ) iff  $B$  is not true in  $L$  (with respect to  $s$ ).
- b. If  $A$  is a negation of a formula  $B$ , then  $A$  is true in  $L$  (with respect to a sequence  $s$ ) iff  $B$  is not true in  $L$  (with respect to  $s$ ).

The resulting abstraction extends the generality of the truth definition to classes of first-order languages that differ arbitrarily in syntax, plus logical and nonlogical vocabulary.

Although this generality is appealing, it has a price. Whereas the original definitions simply stipulated that  $\lceil \neg A \rceil$  is a negation,  $\lceil A \vee B \rceil$  is a disjunction, and  $\lceil \exists x Ax \rceil$  is an existential generalization over a range  $D$  of objects, the revised definition doesn't provide a clue about which formulas fall into these categories. Moreover, Field's physicalist now has to provide reductions of each of these semantic notions.

How might this be done? We are accustomed either to using truth to explain the logical notions or to taking them as primitive, while stipulating that certain symbols are to count as instances of them. Neither of these policies is open to Field. He cannot characterize negation as a symbol that attaches to a formula to form a new formula that is true (with respect to a sequence) iff the original is false (with respect to the sequence); for that would make the reduction of truth to the notions in (13) circular. Nor can he take negation to be primitive and stipulate that  $\lceil \neg s \rceil$  is to be the negation of  $S$ ; for that would fail to give the facts about speakers that explain the semantic properties of  $\lceil \neg s \rceil$ . Although there are alternative approaches, none that I know of is clearly successful.<sup>16</sup> For example, in *The Roots of Reference*<sup>17</sup> Quine attempts to characterize truth-functional operators in terms of community-wide dispositions to assent and dissent. He ends up concluding that indeterminacy between classical and intuitionist construals of the connectives is inevitable. Although I do not accept Quine's argument for this,<sup>18</sup> I do think that the task confronting Field's physicalist is nontrivial. The problems involved in reducing primitive denotation to physical facts are hard enough; adding the logical notions makes the job that much harder.

As I have stressed, the source of this difficulty is the demand that semantic facts be supervenient on physical facts about speakers. In effect, this demand limits adequate definitions to those which legitimate substitution for semantic notions in contexts like (15) and (16).

- (15) If  $L$ -speakers had behaved differently (or been differently constituted), then ' $b$ ' wouldn't have referred (in  $L$ ) to Boston, and ' $C$ ' wouldn't have applied (in  $L$ ) to cities, and ' $Cb \vee Ca$ ' wouldn't have been true (in  $L$ ) iff Boston was a city or Arizona was a city.
- (16) The fact that  $L$ -speakers behave as they do (and are constituted as they are) explains why ' $b$ ' refers (in  $L$ ) to Boston, etc.

Field's critique of Tarski is based on the conviction that there ought to be a way of spelling out (15) and (16) so that they come out true when physicalistic substitutes replace semantic terms and their initial clauses are construed as expressing contingent physical possibilities.<sup>19</sup> As we have seen, Tarski's definition doesn't have this character.

### III

It is helpful in understanding the issues at stake to compare this criticism of Tarski to a parallel objection. Whereas Field's critique is based upon a view about the relationship between speakers and semantic properties like truth, the parallel objection is based on a view about the relationship between meaning and truth. It is widely held that the meaning of a sentence is closely related to its truth conditions and that knowledge of the one constrains knowledge of the other. Thus, many philosophers would accept arbitrary instances of (17) and (18):

- (17) If 'S' had meant in *L* that *p*, then 'S' would have been true in *L* iff *p*.
- (18) If *x* knows that it is not the case that 'S' is true in *L* iff *p*, then *x* knows (or has sufficient grounds for concluding) that 'S' does not mean in *L* that *p*.<sup>20</sup>

A natural demand growing out of this view is that substituting an adequate explication for 'true in *L*' in (17) and (18) should result in true sentences with contingent antecedents.<sup>21</sup>

As before, it is obvious that Tarski's definition does not satisfy this demand. For example, let '*Ws*' be a sentence of *L* meaning that snow is white. Using Tarski's definition of truth, we can produce the following counterparts of (17) and (18):<sup>22</sup>

- (17<sub>T</sub>) If '*Ws*' had meant in *L* that snow is black, then it would have been the case that snow was white iff snow was black.
- (18<sub>T</sub>) If *x* knows that it is not the case that snow is white iff snow is black, then *x* knows (or has sufficient grounds for concluding) that '*Ws*' does not mean in *L* that snow is black.

These are clearly not what the defender of (17) and (18) has in mind. The reason they aren't is that Tarski's set-theoretic truth predicate doesn't impose any conditions on the meanings of the sentences to which it applies. To be sure, Tarski wouldn't count any predicate  $T$  as a truth predicate unless  $\ulcorner \alpha \text{ is } T \urcorner$  were materially equivalent to any metalanguage paraphrase of the object-language sentence named by  $\alpha$ . On the basis of this, one might interpret Tarski as implicitly supposing that instances of (19) are necessary or a priori.

- (19) If ' $T$ ' is a truth predicate for  $L$ , and ' $S$ ' means in  $L$  that  $p$ , then ' $S$ ' is  $T$  iff  $p$ .

However, this is quite different from maintaining that if ' $T$ ' in (20) is replaced with a truth predicate for  $L$ , then the resulting instances of the schema will be necessary or a priori:

- (20) If ' $S$ ' means in  $L$  that  $p$ , then ' $S$ ' is  $T$  iff  $p$ .

It is this that the advocate of (17) and (18) demands and that Tarski appears not to provide.<sup>23</sup>

#### IV

We have, then, two major objections to Tarski. Field demands that semantic properties be dependent on speakers in a way in which Tarski's substitutes are not. A familiar sort of semantic theorist demands that meaning and truth conditions be contingent, but analytically connected, properties of a sentence in a manner incompatible with Tarski. The only way to defend Tarski's philosophical interpretation of his work is to reject these demands.

Although this might initially seem to be a desperate strategy, it is not. Think of a standard first-order language  $L$  as a triple  $\langle S_L, D_L, F_L \rangle$ , where  $S_L$  is a family of sets representing the various categories of well-formed expressions of  $L$ ;  $D_L$  is a domain of objects; and  $F_L$  is a function that assigns objects in  $D_L$  to the names of  $L$ , subsets of the domain to one-place predicates of  $L$ , and so on.<sup>24</sup> Let  $J$  be a class of such languages. Truth can now be defined in nonsemantic terms for variable ' $L$ ' in  $J$  in a straightforward Tarskian fashion. The only significant change from before is that the notions of primitive denotation are no longer given language-

specific list definitions, but rather are defined for variable '*L*' using the "interpretation" functions built into the languages. In particular, a name *n* refers to an object *o* in a language *L* iff  $F_L(n) = o$ .<sup>25</sup> The resulting truth predicate is just what is needed for metatheoretical studies of the nature, structure, and scope of a wide variety of theories.

What the truth definition does not do is tell us anything about the speakers of the languages to which it applies. On this conception, languages are abstract objects, which can be thought of as bearing their semantic properties essentially. There is no possibility that expressions of a language might have denoted something other than what they do denote; or that the sentences of a language might have had different truth conditions. Any variation in semantic properties (across worlds) is a variation in languages. Thus, semantic properties aren't contingent on anything, let alone speaker behavior.

What is contingent on speaker behavior is which language a person or population speaks and which expression a given utterance is an utterance of. Let  $L_1$  and  $L_2$  be two languages in *J* which are identical except for the interpretations of certain nonlogical vocabulary—perhaps the color words in  $L_1$  are shape words in  $L_2$ . We can easily imagine a situation in which it is correct to characterize  $L_1$ , rather than  $L_2$ , as the language of a given population. To ask what such a characterization amounts to, and what would justify it, is to ask not a semantic question about the languages, but a pragmatic question about their relation to speakers.

Although Tarski had nothing to say about this relation, other philosophers have. David Lewis, using a different, but equally abstract, conception of language has proposed (op. cit.) an analysis in terms of a convention of truthfulness and trust. Discussions of what Donald Davidson calls "radical interpretation" can also be reconstructed as dealing with this issue. For physicalists, the interesting question is whether any purely physical explication can be given. If so, then the physicalist can accept both semantic notions that apply to sentences and those which apply to utterances. If not, then either the latter or physicalism itself must be abandoned.

It is interesting to note that much of Hartry Field's concern is with the semantic properties of utterances rather than sentences. In describing the physicalist's position he says:

People utter the sounds ‘Electrons have rest mass but photons don’t’ . . . , and we apply the word ‘true’ to their utterances. We don’t want to say that it is a primitive and inexplicable fact about those utterances that they are true, a fact that cannot be explicated in non-semantic terms; this is as unattractive to a physicalist as supposing that it is a primitive and inexplicable fact about an organism at a certain time that it is in pain. (op. cit., P. 378 herein)

In effect, Field criticizes Tarski for not providing a physicalistically acceptable truth predicate of utterances. But Tarski wasn’t concerned with utterances. Thus, confronted with the question

(i) In virtue of what are certain sounds utterances which are true in *L*?

Tarski’s response ought to be to break it up into two subsidiary questions:

(ii) In virtue of what are certain sounds utterances in *L* of its sentences?

(iii) In virtue of what are sentences of *L* true (in *L*)?

Whereas Tarski answered the second question, the first was not part of his task.

It is hard to see how Field himself could avoid this division of labor. At one point he suggests that in order to handle ambiguous and indexical expressions, truth definitions should be formulated in terms of tokens rather than types (pp. 369–371 herein). The idea is that utterances are contextually disambiguated and that semantic notions should apply to unambiguous entities. This means that all clauses in a truth definition must be formulated as applying to tokens. To this end, Field reformulates the clause for negation as (21):

(21) A token of  $\neg e$  is true (with respect to a sequence) iff the token of *e* that it contains is not true (with respect to the sequence).  
(P. 370 herein)

However, this won’t do. As I indicated earlier, Field can’t accept any truth definition in which a certain syntactic form is simply stipulated to be a negation; for to do this would be to fail to explicate the facts about speakers in virtue of which negative constructions have the semantic properties they do. Instead, (21) must be replaced with something along the lines of (22).

- (22) A token of a formula *A*, which is a negation of a formula *B*, is true (with respect to a sequence) iff some designate token of *B* is not true (with respect to the sequence).

But now there is a problem. Even if the notion of a formula *A* being a negation of a formula *B* can be given a physicalistic definition in terms of the behavior of speakers, there is no clear way of specifying the relevant token of *B* needed in (22): indeed, there is no way of ensuring that it will exist.

If we could count on utterances of negative sentences always containing, as proper parts, utterances of the sentences they are negations of, then the problem would not arise. Although this is a feature of certain artificial languages, it is not a characteristic of natural languages actually spoken by people. In order to avoid arbitrarily restricting truth definitions to (utterances involving) this subset of artificial languages, we need some way of eliminating undue dependence on empirically unreliable tokens. The most straightforward way of doing this is to define truth for types, thereby acknowledging the theoretical division of labor I have attributed to Tarski.<sup>26</sup> Once this is done, the physicalist is free to accept Tarski-like truth definitions applying to sentences, while leaving it open whether the pragmatic relations between languages, expressions, speakers, and utterances are purely physicalistic.<sup>27</sup>

It should be emphasized that although the linguistic threat to physicalism has been moved from semantics to pragmatics, it is still a serious one. It is by no means evident that physicalistic reductions of the crucial relations can be given. One physicalist who seems to think they cannot be given is Quine. Although he doesn't conceptualize matters in just the way that I have, it is illuminating to interpret him as accepting Tarski's semantic definitions while rejecting any physicalistic reduction of the pragmatic notions. On this interpretation, there is no indeterminacy about the claim that 'rabbit' refers to rabbits in a certain specified language, call it "English," or about the claim that 'gavagai' refers to rabbits in another language, call it "Junglese." What is indeterminate is whether I speak English, as opposed to some related rabbit-stage language, and whether the native speaks Junglese, as opposed to some similar counterpart.

The upshot of this is that it is all right for a Quinean physicalist to *use* a Tarskian language to describe the world, and even to attribute Tarskian

semantic properties to expressions in that language. What he cannot do is identify the language he is using. When it comes to describing linguistic behavior—even one's own—identifiable Tarskian languages are excluded in favor of dispositions to verbal behavior. The strain in this position is a measure of the challenge that language use presents to physicalism. What is not problematic is the physicalist's acceptance of Tarski.

## V

This discussion illustrates a general strategy for answering Tarski's critics. Field's objection was that Tarski's semantic properties are not dependent on facts about speakers. The Tarskian reply is that nothing is lost by thinking of semantics abstractly and relegating the interpretation of speakers' behavior to pragmatics. In so doing, one gains the advantages of a truth predicate for metatheoretical discussions, while retaining the ability to raise deep philosophical problems in other areas.

As I pointed out earlier, Field's is not the only objection to Tarski. Any theory of semantic competence that makes knowledge of truth conditions the central notion implicitly rejects Tarski's claim to have provided a notion of truth adequate for all theoretical purposes. The defense against this objection is that such theories are flawed in any case.

The problem with these theories lies in specifying what truth conditions are in such a way that knowledge of them is necessary and sufficient for understanding. If we assume that truth conditions involve the notion of truth, then it is natural to suppose that they are given by T-sentences of the form (23):

(23) 'S' is true in  $L \equiv P$

(Instances are formed by replacing 'P' with a sentence that means the same as the sentence replacing 'S'.) However, it is easy to show that knowing the propositions expressed by T-sentences is neither necessary nor sufficient for understanding meaning (where 'true' is taken to be a non-Tarskian primitive and ' $\equiv$ ' represents either material or necessary equivalence). Thus it is not obvious that what one knows when one understands a language involves the notion of truth at all. If it doesn't, it may be that nothing is lost by adopting a Tarski-like explication of truth together with an independent account of semantic competence.



Although I won't try to show it here, I think that this is the right approach for both truth and semantic competence. This does not mean that Tarski's semantic predicates really are adequate for all theoretical purposes. Saul Kripke's theory of truth is a genuine advance on Tarski's treatment of the liar.<sup>28</sup> In addition, semantic predicates for richer languages, as well as for propositions, are needed. What does seem right about Tarski's approach is its deflationist character. Theories of truth for sentence types need not specify the facts about speakers in virtue of which their utterances have content; nor should such theories be seen as issuing in theorems knowledge of which is necessary and sufficient for semantic competence. Instead, theories, or definitions, of truth should provide accounts of the content of familiar truth predications, while resolving the semantic paradoxes (and their propositional variants). Beyond this, and the attendant dissolution of confusions, it is best not to expect too much. Truth is a useful notion, but it is not the key to what there is, or to how we represent the world to ourselves through language.

#### Notes

This paper was presented in the spring of 1983 at Yale and Dartmouth, where I profited from useful discussion. Thanks are also due to Paul Benacerraf, Axel Buhler, John Burgess, Gilbert Harman, David Lewis, and Walter Sinnott-Armstrong for reading and commenting on an earlier draft.

1. Various versions of the "redundancy theory" fall into this category. Although these versions deal with "propositional" contexts like those in (1), they deny that 'true' is predicated of propositions, or anything else. Arguments against these approaches are given in my *Semantic Theories of Truth* (New York: Oxford, in preparation).

2. "The Semantic Conception of Truth," sec. 15, in Leonard Linski, ed., *Semantics and the Philosophy of Language* (Urbana: U. of Illinois Press, 1952). (Reprinted as chap. 15, this volume.—Ed.)

3. *The Philosophy of Logic* (Englewood Cliffs, N. J.: Prentice-Hall, 1970), pp. 10–13.

4. Or, equivalently, (4'):

(4') For all sentences  $p$ ,  $q$   $\vdash (p \rightarrow (q \rightarrow p))^\top$  is true.

5. There are, I presume, many versions of realism, of which (6) and (7) represent only one. A different version might hold that some sentence of English is such that it is metaphysically possible for it to be true (keeping the semantics fixed) in cases in which the proposition it expresses cannot (ever) be known or rationally believed. This thesis is no more linguistic than (6) and (7) are.

6. And also of skepticism about related notions like definability. Tarski cites the paradoxes as a source of skepticism in "The Establishment of Scientific Semantics" and, as John Burgess has pointed out to me, in "On Definable Sets of Real Numbers," reprinted in *Logic, Semantics, Metamathematics* (New York: Oxford, 1956).

7. See C. G. Hempel, "On the Logical Positivists' Theory of Truth," *Analysis*, II, 4 (January 1935): 49–59; Rudolf Carnap, "Intellectual Autobiography," in P. A. Schilpp, ed., *The Philosophy of Rudolf Carnap* (LaSalle Ill.: Open Court, 1963), p. 63; Carnap, "Truth and Confirmation," in H. Feigl and W. Sellars, eds., *Readings in Philosophical Analysis* (New York: Appleton-Century-Crofts, 1949); Hans Reichenbach, *Experience and Prediction* (Chicago: University Press, 1938), sec. 22; Otto Neurath, "Sociology and Physicalism" and "Protocol Sentences", both reprinted in A. J. Ayer, *Logical Positivism* (Glencoe Ill.: Free Press, 1959); Karl Popper, *The Logic of Scientific Discovery* (New York: Harper Torchbook, 1965), p. 274; and Hartry Field, "Tarski's Theory of Truth," *Journal of Philosophy* LXIX, 13 (July 13, 1972): 347–375 (reprinted as chap. 16 in this volume—Ed.); page references to Field will be to this article as reprinted herein.

8. "The Establishment of Scientific Semantics," op. cit., p. 406.

9. Tarski's physicalism countenanced both physical science and "logic," where the latter was construed as including set theory and everything obtainable from it. In what follows, I will use the term 'physicalism' in the moderate sense of (i–iii) above. In particular, physicalism, in my sense, does not require the reduction of set-theoretical facts, mathematical facts, or syntactic facts about expression types.

10. There are, of course, technical issues as well. When the metalanguage contains quantifiers ranging over arbitrary subsets of the domain of the object language, an explicit definition of object-language truth is possible in the metalanguage. On the other hand, if a classical object language containing set theory has quantifiers ranging over all sets, then an explicit metalanguage definition of truth is impossible. Tarski's emphasis on explicit definition is philosophical in the sense that he saw significant philosophical advantages in explicit definitions of truth, *where they are possible*.

11. Note, since there are only finitely many atomic formulas in  $L$ , that we could have got an equivalent result by substituting (i') for (i) in (T).

(i')  $x = 'Ca'$  and Arizona is a city, or  $x = 'Cb'$  and Boston is a city, or  $x = 'Sa'$  and Arizona is a state, or  $x = 'Sb'$  and Boston is a state, ... (and so on for each atomic formula).

12. I use the phrase 'linguistic behavior' in a broad sense to include all facts about speakers relating to their use of language.

13. Field also includes the notion of a function sign being fulfilled by a pair of objects. In the interest of simplicity, I am ignoring this.

14. Presumably, speakers of  $L_1$  differ in some way from speakers of  $L_2$  regarding their beliefs, intentions, attitudes, brain states, or conditioned responses involving 'v'.

15. Field partially anticipates this point in footnotes 5 and 10 of his paper. In fn 5 he notes that in model theory quantifiers are given an “unusual” semantics in which they range over the members of some specified set, rather than over all (actually existing) things. In such a case, Field claims, Tarski has reduced truth to primitive denotation, plus the notion of the range of the quantifiers. (For Tarski this constituted the usual case, since it is only when the range of quantifiers is restricted that explicit truth definitions can be given—for languages with a certain minimal richness.)

In fn 10 Field notes, without specifying, the existence of problems that must be faced when the definition of truth is generalized so as not to contain a particular logical vocabulary.

16. The most interesting, in my opinion, is briefly sketched in Gilbert Harman, “Beliefs and Concepts,” *Proceedings of the Philosophy of Science Association*, II (1982).

17. LaSalle, Ill.: Open Court, 1974.

18. For an excellent critique of Quine, see Alan Berger, “Quine on ‘Alternative Logics’ and Verdict Tables,” *Journal of Philosophy* LXXVII, 5 (May 1980): 259–277.

19. In stating this requirement in terms of the replacement of a semantic term by its physicalistic definiens, I have tacitly relied on Tarski’s emphasis on explicit definition. However, I don’t think the philosophical point of the requirement depends on this. In cases in which only an axiomatic treatment is possible, Field could require that the axioms governing ‘true’, together with empirical facts about speakers and their environments, have statements of type (15) and (16) as consequences.

20. (18) is considerably weaker than the claim that knowledge of truth conditions is sufficient for knowledge of meaning. (18) says only that knowledge of truth conditions is capable of providing some information about meaning. In effect, it says that even if knowledge that

(i) ‘S’ is true in *L* iff *q*

is not sufficient for knowing that

(ii) ‘S’ means in *L* that *q*

it should be sufficient for knowing that

(iii) ‘S’ does not mean in *L* that *p*

(where the sentences replacing ‘*p*’ and ‘*q*’ are obviously incompatible).

21. Although the contexts in question are intensional, this demand does not require that an adequate explicatum for the pretheoretic notion of truth be intensionally equivalent to an ordinary, pretheoretic truth predicate. Rather, it requires that all legitimate theoretical purposes served by the explicandum (truth), be equally well served by the explicatum. For example, if knowledge of that expressed by

(i) ... is true ...

is used to help explain the nature of some capacity (say, the capacity to understand sentences), then knowledge of that expressed by

(ii) ... T ...

(where the explicatum T is substituted for 'is true') should be sufficient for the same purpose. An explication that meets this requirement of theoretical productivity will allow the explicandum to be eliminated from one's total scientific and philosophical theory without loss of explanatory power. Thus, substitution of explicatum for explicandum in intensional contexts contained in one's total explanatory theory must be countenanced, even if such substitution is not always countenanced in ordinary discourse.

The qualification in fn 19 above regarding substitution and explicit definition also applies here.

22. (17<sub>T</sub>) and (18<sub>T</sub>) are simplifications of the sentences that would result from substituting Tarski's explicatum [the right-hand side of (T') in section II] for 'true in L' in (17) and (18). The simplifications are based on the fact that, where T is Tarski's explicatum, 'Snow is white' is T and 'Snow is white' are necessarily equivalent (in the presence of elementary set theory). In light of this equivalence, replacing one with the other should not affect the philosophical issues at stake in (17) and (18).

23. Hilary Putnam has used a version of the argument involving (17)/(17<sub>T</sub>) against Tarski (in a lecture at Princeton, fall 1982). Michael Dummett has used a version of the argument involving (18)/(18<sub>T</sub>) against Tarski (in the preface to *Truth and Other Enigmas* [Cambridge, Mass.: Harvard, 1978], and in "Truth," reprinted there).

The arguments given above are intended as stand-ins for a variety of related arguments, all designed to show that Tarski's notion of truth has nothing to do with semantic interpretation or understanding. For example, it is probably best to understand Davidson not as attempting to analyze meaning in terms of truth, but rather as eliminating the notion of meaning in favor of the notion of truth. Since (18) utilizes the notion of meaning, a defender of the Davidson of "Truth and Meaning" might want to trade it for something like (i):

- (i) If  $x$  knows that which is expressed by the relevant instance of  
       'S' is true in  $L$  iff  $p$   
       for each sentence of  $L$ , then  $x$  is a competent speaker of  $L$ .

If 'true in  $L$ ' is understood as short for the definiens provided by Tarski, (i) is as absurd as (18<sub>T</sub>).

Just this sort of absurdity is present in familiar and often repeated remarks like the following (which would allow Tarski's definiens to be the central notion in a theory of meaning):

(T)  $s$  is  $T$  if and only if  $p$

What we require of a theory of meaning for a language  $L$  is that without appeal to any (further) semantical notions it place enough restrictions on the predicate 'is  $T$ ' to entail all sentences got from schema T when ' $s$ ' is replaced by a structural description of a sentence of  $L$  and ' $p$ ' by that sentence....

It is worth emphasizing that the concept of truth played no ostensible role in stating our original problem. That problem, upon refinement, led to the view that an adequate theory of meaning must characterize a predicate meeting certain conditions. It was in the nature of a discovery that such a predicate would apply exactly to the true sentences. I hope that what I am doing may be described in part as defending the philosophical importance of Tarski's semantical concept of truth. (Donald Davidson, "Truth and Meaning," in J. F. Rosenberg and C. Travis, eds., *Readings in the Philosophy of Language* [Englewood Cliffs, N.J.: Prentice-Hall, 1971], pp. 455/56.)

Earlier statements of essentially the same absurdity can be found in Rudolf Carnap's *Meaning and Necessity* (Chicago: University Press, 1947), pp. 5/6; and in section 7 of his *Introduction to Semantics* (Cambridge, Mass.: Harvard, 1943).

24. This sort of construction is familiar from model theory. However, its use here is different from model-theoretic treatments. Here we are not defining *truth in L relative to a model*, but rather *truth in L* (simpliciter) for an enriched conception of a language. This way of looking at things was suggested to me from two sources: David Lewis's "Languages and Language," in K. Gunderson, ed., *Minnesota Studies in the Philosophy of Science*, VII (Minneapolis: Univ. of Minnesota Press, 1975), pp. 3–35; and one of Saul Kripke's seminars on truth, Princeton, 1982.

25. Note,  $F_L$  is a purely mathematical object—a set of pairs, if you like. Thus, it does not incorporate any undefined semantic notions. This was one of the points noted by Kripke in the seminar mentioned in fn 24.

26. Ambiguity can then be treated as a case of homonymy. For example, instead of thinking that English contains a single (ambiguous) word type 'bank', one can take English to contain two different words, 'bank<sub>1</sub>' and 'bank<sub>2</sub>', whose tokens are phonologically identical. The contextual factors that Field relies on to disambiguate tokens can then be thought of as determining whether particular utterances are tokens of the type 'bank<sub>1</sub>' or the type 'bank<sub>2</sub>'.

27. Acknowledging the need to formulate truth definitions in terms of types does not force one to think of the semantic properties of sentences as invariant from world to world and not dependent on the properties of tokens. However, it does make this a natural position.

28. "Outline of a Theory of Truth," *Journal of Philosophy* LXXII, 19 (Nov. 6, 1975): 690–716.

# VI

---

## Deflationary Views and Their Critics



---

# Deflationary Views and Their Critics:

## Introduction

“Deflationism” is a name for a family of views, all of which aim to deflate the lofty pretensions of traditional theories of truth. Correspondence or coherence theories aim to tell us about the inner nature of truth, but deflationists believe that truth has no nature. As a consequence, they hold that truth is a less important concept than other philosophers have led us to think. The roots of deflationism are in Ramsey 1927 and chap. 18, Frege 1956, and possibly Tarski, chap. 15, but the view really began to take the philosophical world by storm after the second world war, spurred in large part by the seminal work of Quine, (chap. 20) and Strawson (chap. 19).

The case for any sort of deflationism generally starts by calling attention to the so-called transparency of truth. When we consider that it is true that roses are red, it seems that we can “look right through” its truth and simply consider that roses are red. In other words, we automatically infer that it is true that roses are red from the proposition that roses are red, and vice versa. So in general, we have the following:

(TS) The (belief, utterance, proposition, etc.) that  $p$  is true if and only if  $p$ .

As Paul Horwich notes (chap. 24), the basic intuition behind deflationism is that there is nothing more to our concept of truth than our commitment to instances of (TS). In the deflationist view, there is no reason to try to explain why (TS) is true—by appealing to deep facts about correspondence or coherence, for instance—(TS) itself (more or less, depending on the variety of deflationism involved) explains all there is to explain about our concept of truth.



Any deflationary view of truth is comprised of two basic components. The first is a *metaphysical* thesis: truth has no nature. This is often put by saying that “true” does not express a property, or at least no real or substantive property. The second component is *semantic*. Since “true” does not pick out a property that all true statements share, the deflationist must explain the meaning of the word “true” and/or the purpose of having such a word or concept in our language. As the essays in this part reveal, various deflationary positions spell out these aspects of the position in different ways.

### The Classical Redundancy Theory

The simplest form of deflationism is the so-called “redundancy theory” often credited to Frank Ramsey (1927, chap. 18). According to Ramsey, “it is really obvious” that “The earth is round” is equivalent to “It is true that the earth is round.” When ascribing truth to a single proposition, he remarks elsewhere, “is true” is “an obviously superfluous addition” to the original assertion (1927, 37). Therefore, we only *appear* to ascribe a property to a sentence or proposition when we say that it is true; we are in reality ascribing nothing, saying nothing more than if we had simply stated the proposition itself.

It is often thought that Ramsey believed that *all* ascriptions of truth are similarly gratuitous or “redundant.” And indeed, in the essay included here, he endorses what looks like a general definition: “A belief that *p*, we say, is true if and only if *p*.” Yet Ramsey was well aware (1927, 36) that any generalized redundancy account faces problems with *blind ascriptions of truth*, such as “Everything she said was true.” For “Everything she said was true” is clearly not equivalent to everything she said. Nor can the simple redundancy view explain generalizations involving truth, such as the statement that every proposition is either true or false. Without a satisfactory explanation of such uses of “true,” no generalized redundancy theory is plausible.

Ramsey was also aware that it was difficult to even state the redundancy theory. For instance, the natural suggestion (R) faces serious problems:<sup>1</sup>

(R)  $x$  is a true belief iff  $(\exists x)(x = \text{the belief that } p \ \& \ p)$

As he notes, the phrase “‘and  $p$ ’ sounds [like] nonsense because it seems to have no verb.” The problem, in short, is that if we interpret the quantifier objectually, “ $x = \text{the belief that } p \ \& \ p$ ” is not grammatically formed; “ $p$ ” can’t serve as a conjunct here.<sup>2</sup> Finally, the question arises as to why we would have the word “true” in our language if it is gratuitous, as the generalized redundancy account suggests. Under the plausible assumption that the word “true” does have a function in our language, the redundancy theory can’t seem to say what it is.<sup>3</sup>

### Strawson’s Performative Theory

According to P. F. Strawson, apparent ascriptions of truth to statements or propositions are actually nonassertoric performative utterances. An example of a nonassertoric performative utterance would be a command. If I tell you to close the door, I am not making an assertion, or stating a fact. I am telling you to do something. Strawson argues that we should regard utterances of the form “It is true that  $p$ ” in a similar way. When I say that it is true that the door is open, I am not describing but endorsing, or agreeing to the claim that the door is open. We use the truth predicate not to ascribe a property but to endorse.

The performative theory faces some formidable objections. First, the word “true” seems to have other uses than to act as a device of endorsement (e.g., to concede a point: “What you say is true but I still think there is something fishy going on”; to connect inferences: “If it is true that  $p$ , then  $q$ ”; etc.) Second, the fact that “it is true that” can express agreement needn’t rule out (as Strawson conceded, 1964) that it can *also* be used to say something about a particular proposition, namely, that it has the property of being true. Thus the theory seems compatible with nondeflationary views of truth as well.

Strawson’s theory does explain what the redundancy theory could not, namely, what the function of “true” is in our language. But the performative theory is also important because it calls our attention to an often neglected feature of our concept of truth: its normative role in our language. Part of our concept of truth is that having true beliefs is valuable. Any deflationary view must explain the normative force of the concept, and it must do so without relying on robust metaphysical facts

about the nature of truth. As a result, it seems that either Strawson's performative theory or a variant will always be part of any deflationary account of truth.

### Disquotationalism

Another classic deflationist view is disquotationalism. Disquotationalism is so called because of Quine's remark, "To ascribe truth to ['Snow is white'] is to ascribe whiteness to snow," and thus "Ascription of truth just cancels the quotation marks. Truth is disquotation" (chap. 20).

Disquotationalism is therefore a cousin of the redundancy theory. Yet disquotationalism differs from the redundancy theory in several important respects. First, it treats sentences, rather than propositions or statements, as the primary bearers of truth and falsity.<sup>4</sup> This sits well with those philosophers, most notably Quine himself, who distrust talk of sentence-shaped but nonlinguistic "propositions." Second, disquotationalism explicitly acknowledges that truth has a crucial function within our language (see Leeds 1979). In the disquotationalist view, the truth predicate is still superfluous when assigned to a particular sentence. But unlike the redundancy theory, the disquotationalist acknowledges that we need the word "true" to generalize over a potentially infinite number of sentences, as we do when we say "Everything Socrates said was true" or "All sentences of the form 'If  $p$  then  $p$ ' are true." In predicating truth to a sentence of the form "If  $p$  then  $p$ ," the truth predicate acts as a device for expressing an infinite conjunction of sentences of that form:

(If roses are red then roses are red) and  
 (If violets are blue then violets are blue) and  
 ⋮

Further, disquotationalists argue that the truth predicate plays this expressive role *because of* its disquotational nature (see, e.g., Field, chap. 21). Thus the disquotationalist turns a powerful objection against the redundancy theory into a virtue of disquotationalism. The importance of the truth predicate is that it provides us with a way to express infinite conjunctions and disjunctions, just as we can express "Snow is white" by calling that sentence true. As such, "true" in the disquotational sense

gives us a usefully indirect way to speak of the world. So, once we have explained its disquotational function and shown why it is useful to have such a predicate, we have said all that needs saying about truth.<sup>5</sup>

This picture is not without its problems. According to the disquotationalist, the truth predicate allows us to express infinite conjunctions and disjunctions, just as we can express “Snow is white” by calling that sentence true. Anil Gupta claims (chap. 23) that this point only makes sense if “express” here is interpreted as “means the same as,” yet asserting a generalization is not the same as asserting its instances. Second, disquotationalists still face the question of how to state their theory. One method defines “ $x$  is true” in such a way that it is simply equivalent to an infinite disjunction of sentences, as in (D):

(D)  $x$  is a true sentence iff  $(x = a \text{ and } a)$  or  $(x = b \text{ and } b)$  or ...

But this raises the disturbing question of how any human being could ever have a complete grasp of the meaning of “true.” According to the disquotationalist, every disjunct on the right hand side of (D) partially defines “ $x$  is true.” But as Gupta notes, this list is infinitely long. I cannot grasp it all. So I cannot grasp a complete definition of truth.

In his contribution to this part, Field argues that disquotationalists should restrict their view so that “true” applies only to those sentences the speaker understands. This implies that we don’t understand the claim that a sentence is disquotationally true any more than we understand the sentence itself. Yet Field’s suggestion ensures that I *do* understand (D), for every sentence on the right-hand side will necessarily be one that I actually understand. But he remains mute on how truth might work for sentences I don’t understand, and intuitively, some sentences that I don’t understand are true.

### The Prosentential Theory

All of us can use pronouns, as in “Bridget liked the play because *she* thought *it* was funny.” Here “she” and “it” operate anaphorically by picking up the reference of their antecedents. In their influential paper “A Prosentential Theory of Truth” (1975), Grover, Camp, and Belnap proposed that there are also *prosentences*. Prosentences are to sentences

as pronouns are to nouns.<sup>6</sup> A familiar example of a prosentence is the word “so,” e.g., “Tom said that Bridget liked comedies; if so, she should go to the play with us.” It is the chief contention of the prosentential theory of truth that “It is true” and “That’s true” function in the same way. The essential idea can be expressed by way of an analogy with an imaginary atomic prosentence “thatt,” as in the following dialogue:

*Tom* Snow is white.

*Bridget* Thatt.

The idea behind the prosentential theory is that the deep structure of “That is true” is analogous to the atomic prosentence “thatt.” Thus the individual words in either “It is true” or “That is true” are separable components only in the grammatical sense; as parts of a prosentence, they carry no individual content. Further, Grover claims that all other uses of “true” are either redundant or can be analyzed prosententially. As a result, the advertised advantage of this theory is that while it maintains that prosentential truth has a role to play in our language, we can eliminate “true” as a separable *predicate*, thus freeing the philosopher from explaining the nature of the property that such a predicate would express.

In a new essay included here, Dorothy Grover argues that one significant advantage that prosentential theories of truth have over other deflationary positions is that the prosententialist does not see truth primarily as a means of “semantic ascent.” Rather, the prosentential truth predicate typically works at the level of the object language. In Grover’s view, this aspect of prosententialism allows the prosententialist to say that truth has, in a sense, an important role to play in explanation and linguistic understanding.

### Horwich’s Minimalist Theory

Like disquotationalism, Paul Horwich’s minimalism takes the concept of truth to function as a device for generalizing over unstated and infinite numbers of propositions. But unlike the disquotationalist or prosentential theories, minimalism takes propositions as the primary truth bearers, where propositions are the contents of our beliefs and utterances, or

*what* is believed or uttered. Thus the minimal theory has no qualms about applying truth to sentences we don't understand. Further, Horwich explicitly eschews any attempt to define or provide an analysis of truth. Rather, the minimal theory (MT) simply takes as axioms every instance of (E):

(E) The proposition that  $p$  is true iff  $p$ .

By stipulating that the axioms of MT are comprised of every instance of this propositional form, Horwich sidesteps the problem we encountered above of generalizing a deflationary theory. But note that MT will therefore have an infinite number of axioms. Not only will there be an infinite number of axioms that we could formulate in current English, there will also be an infinite number of axioms that we cannot yet formulate in current English. This might seem to imply that we can never grasp the concept of truth. But Horwich argues that what constitutes our grasp of the concept of truth is our *inclination* to accept instances of (E) that we would understand, not our actual acceptance of them. In Horwich's view, this inclination is "an explanatorily basic regularity," or a fact that cannot be explained in terms of other facts and that in turn explains every other fact about our use of the word "true" (1998, 34). Our understanding of "true" is in this sense basic. Finally, Horwich is more willing than other deflationists to accept that truth is a property, since "true" is a predicate and in Horwich's view, all predicates express properties in a minimal sense. Yet Horwich remains a deflationist, since he holds (a) that truth is a simple logical property with no underlying nature, and (b) that truth is not a property that does any explanatory work.

This last claim is vigorously challenged by Donald Davidson in his essay in this volume (chap. 26, included in part VII). Davidson argues that truth is an explanatory concept because we need to appeal to facts about truth in order to explain other important philosophical concepts. According to Davidson's influential theory of meaning, to know the meaning of a sentence is to know the conditions under which the sentence is true. If this is so, then truth is needed to explain meaning. Anil Gupta (chap. 23) also presses Horwich on whether his theory has the resources to explain how we know certain general facts about truth. For while our commitment to accepting instances of (E) and logic can explain why we accept

particular instance of “It is true that if  $p$ , then  $p$ ,” it can’t seem to explain our acceptance of the generalization that it is true that every proposition implies itself. In the new essay included here, Horwich mounts an ingenious defense of his position against these and other objections by Gupta (chap. 23), Davidson (chap. 26), Dummett (chap. 10), Soames (chap. 17), and Field (chap. 21).

### Deflationism and Correspondence

A remaining question is the relationship between deflationism and other views of truth. Hartry Field takes a truly deflationary theory to hold that the only useful view of truth is a disquotational notion like that expressed by (D) above. But (D) itself Field takes as neutral between deflationary and robust theories of truth. Thus he suggests that correspondence theories of truth are best seen as *extensions* of disquotational theories. In his view, the difference between the two consists in the fact that in the correspondence view, our utterances have, in addition to purely disquotational truth conditions, “correspondence truth conditions,” which are “objective features of an utterance or thought-state, features which the utterance or thought-state could have whether or not we know of it.” As a consequence, correspondence theories aren’t inconsistent with disquotationalism; rather, like any robust theory of truth, they require the disquotational notion so that we can generalize over a potentially infinite number of propositions, as we do when we say “Everything you say is true.”

In his selection, Michael Devitt contrasts deflationism with the correspondence theory. Recall that any deflationary position has two aspects: a metaphysical aspect (truth has no nature) and a semantic aspect (the role or purpose of the word “true”). Devitt argues that the metaphysical component of any sort of deflationism entails that it must be seen as a type of antirealism about truth. Unlike views such as the correspondence theory, deflationists can’t, strictly speaking, believe that there is any “reality to truth.” Second, since deflationism entails that the predicate “true” does not serve, like normal predicates, to ascribe a robust property, Devitt argues that the semantic aspect of any deflationary view ultimately revises our ordinary linguistic practice. In short, Devitt sees deflationism

as analogous to “nonfactualist” views in ethics, which hold that statements like “*x* is wrong” do not state facts about the world but instead act, e.g., to express our approval or disapproval of certain acts. This contrasts sharply with the correspondence theory, which, Devitt argues, usefully preserves our ordinary semantic practice regarding “true.”

## Notes

1. In this essay included here, Ramsey says that beliefs are the primary truth bearers, but he individuates these in teams of what he calls their propositional reference.
2. One suggestion would be to employ substitutional quantification to this end. But this is problematic in its own right. For instance, one typically employs the concept of truth in order to explain substitutional quantification, not the other way around. See David 1994 (pp. 83–94) for an excellent discussion of these issues.
3. Since Ramsey was aware of several of these problems, it is not entirely clear that he actually held a complete redundancy view (see Kirkham, 1992, 318–319). Indeed, in his contribution to this part Field (chap. 21) argues that Ramsey is better seen as a type of correspondence theorist.
4. In chap. 20, Quine takes what he calls “eternal sentences,” or sentence types as they would be if they were free from the problems caused by tenses, ambiguity, vagueness, indexicals, and the like, as the truth bearers, but he admits that this is a “refinement” of ordinary language, which, he says, treats sentence tokens as the bearers of truth.
5. Michael Williams (1986, 424) presents deflationism in precisely this way. See also Williams 1999 for an excellent discussion of objections raised against disquotationalism by Davidson, chap. 26.
6. Ramsey foreshadows the theoretical development of prosentences in a remark in his contribution (chap. 18).

## Further Reading for Part VI

The importance of deflationary theories within the context of the current debate means that many of the essays in this volume, including many outside of the present part, contain important reflections upon, and objections to, deflationism. See the essays by Davidson (chap. 26), Dummett (chap. 10), Alston (chap. 3), Putnam (chap. 30), Wright (chap. 32), Sosa (chap. 27), and Rorty (chap. 12).

Ayer, A. J. 1936. *Language, Truth, and Logic*. London: Dover.

Brandom, R. 1984. “Reference Explained Away.” *Journal of Philosophy* 81: 469–492.



- Brandom, R. 1988. "Pragmatism, Phenomenalism, and Truth Talk." In P. French, T. Uehling, and H. Wettstein, eds., *Midwest Studies in Philosophy*, vol. 12. Minneapolis: University of Minnesota Press.
- David, M. 1994. *Correspondence and Disquotation*. Oxford: Oxford University Press.
- Davidson, D. 1990. "The Structure and Content of Truth." *Journal of Philosophy* 87: 279–328.
- Field, H. 1986. "The Deflationary Conception of Truth." In G. MacDonald and C. Wright, eds., *Fact, Science, and Morality*. Oxford: Blackwell.
- Field, H. 1994. "Deflationist Views of Meaning and Content." *Mind* 103: 249–285.
- Frege, G. 1956. "The Thought." trans. A. Quinton and M. Quinton. *Mind* 65: 289–311.
- Grover, D., Camp, J., and Belnap, N. 1975. "A Prosentential Theory of Truth." *Philosophical Studies* 27: 73–125.
- Gupta, A. 1993. "Minimalism." *Philosophical Perspectives* 7: 359–369.
- Hill, C. 1987. "Rudiments of a Theory of Reference." *Notre Dame Journal of Formal Logic* 28: 200–219.
- Hill, C. 1999. "Truth in the Realm of Thought." *Philosophical Studies* 96: 87–121.
- Horwich, P. 1998. *Truth*. Oxford: Oxford University Press.
- Horwich, P. 1996. "Realism and Truth." *Philosophical Perspectives* 10: 187–197.
- Horwich, P. 1997. "Deflationary Truth and the Problem of Aboutness." In E. Villanueva, ed., *Truth*, pp. 95–106. Atascadero: Ridgeview Publishing.
- Kovach, A. 1997. "Deflationism and the Derivation Game." *Mind* 106: 575–579.
- Leeds, S. 1978. "Theories of Reference and Truth." *Erkenntnis* 13: 111–129.
- O'Leary-Hawthorne, J., and Oppy, G. 1997. "Minimalism and Truth." *Noûs* 31: 170–196.
- Price, H. 1988. *Facts and the Function of Truth*. Oxford: Blackwell.
- Putnam, H. 1991. "Does the Disquotational Theory of Truth Really Solve All Philosophical Problems?" *Metaphilosophy* 22: 1–13.
- Quine, W. V. 1970. *Philosophy of Logic*, pp. 10–13. Englewood Cliffs: Prentice-Hall.
- Ramsey, F. P. 1927. "Facts and Propositions." *Proceedings of the Aristotelian Society*, supp. vol. 7: 153–170.
- Soames, S. 1999. Chap. 7 of *Understanding Truth*. Oxford: Oxford University Press.

- Strawson, P. F. 1964. "A Reply to Mr. Warnock." In G. Pitcher, ed., *Truth*. Englewood Cliffs: Prentice-Hall.
- Warnock, G. 1964. "A Problem about Truth." In G. Pitcher, ed., *Truth*. Englewood Cliffs: Prentice-Hall.
- Williams, M. 1986. "Do We Epistemologists Need a Theory of Truth?" *Philosophical Topics* 14: 223–242.
- Williams, M. 1999. "Meaning and Deflationary Truth." *Journal of Philosophy* 96: 545–565.
- Wright, C. 1992. Chap. 1 of *Truth and Objectivity*. Cambridge: Harvard University Press.



---

## The Nature of Truth

Frank Plumpton Ramsey

### 1 What Is Truth?

What is truth? What character is it that we ascribe to an opinion or a statement when we call it 'true'? This is our first question, but before trying to answer it let us reflect for a moment on what it means. For we must distinguish one question, "what is *truth*?", from the quite different question "what is *true*?" If a man asked what was true, the sort of answer he might hope for would either be as complete an enumeration as possible of all truths, i.e., an encyclopaedia, or else a test or criterion of truth, a method by which he could know a truth from a falsehood. But what we are asking for is neither of these things, but something much more modest; we do not hope to learn an infallible means of distinguishing truth from falsehood but simply to know what it is that this word 'true' means. It is a word which we all understand, but if we try to explain it, we can easily get involved, as the history of philosophy shows, in a maze of confusion.<sup>1</sup>

One source of such confusion must be eliminated straight away; besides the primary meaning in which we apply it to statements or opinions, the word true can also be used in a number of derived and metaphorical senses which it is no part of our problem to discuss. Obscure utterances such as "Beauty is truth, truth beauty" we shall make no attempts to elucidate, and confine ourselves to the plain work-a-day sense in which it is true that Charles I was beheaded and that the earth is round.

First we have to consider to what class of things the epithets 'true' and 'false' are primarily applied, since there are three classes which might be suggested. For we use 'true' and 'false' both of mental states,<sup>2</sup> such as

beliefs, judgments, opinions or conjectures; and also of statements or indicative sentences; and thirdly according to some philosophers we apply these terms to 'propositions', which are the objects of judgments and the meaning of sentences, but themselves neither judgments nor sentences.

According to the philosophers who believe in them, it is these propositions which are true or false in the most fundamental sense, a belief being called true or false by an extension of meaning according as what is believed is a true or a false proposition. But in as much as the existence of such things as these propositions is generally (and to my mind rightly) doubted, it seems best to begin not with them but with the mental states of which they are the supposed objects, and to discuss the terms true and false in their application to these mental states, without committing ourselves before we need to any doubtful hypothesis about the nature of their objects.

The third class consisting of statements or indicative sentences is not a serious rival, for it is evident that the truth and falsity of statements depends on their meaning, that is on what people mean by them, the thoughts and opinions which they are intended to convey. And even if, as some say, judgments are no more than sentences uttered to oneself, the truth of such sentences will still not be more primitive than but simply identical with that of judgments.

Our task, then, is, to elucidate the terms true and false as applied to mental states, and as typical of the states with which we are concerned we may take for the moment beliefs. Now whether or not it is philosophically correct to say that they have propositions as objects, beliefs undoubtedly have a characteristic which I make bold to call *propositional reference*. A belief is necessarily a belief that something or other is so-and-so,<sup>3</sup> for instance that the earth is flat; and it is this aspect of it, its being "that the earth is flat" that I propose to call its propositional reference. So important is this character of propositional reference that we are apt to forget that a belief has any other aspects of characters at all, and when two men both believe that the earth is flat we say they have the same belief, though they may believe it at different times for different reasons and with different degrees of conviction and use different languages or systems of imagery; if the propositional references are the same,

if they are both “beliefs that” the same thing, we commonly ignore all other differences between them and call them the same belief.

It is usual in logic to express this resemblance between the two men’s beliefs not by saying as I do that they have the same propositional reference, but by calling them beliefs in the same proposition; to say this is not however to deny the existence of such a character as propositional reference, but merely to put forward a certain view as to how this character should be analysed. For no one can deny that in speaking of a belief as a belief that the earth is flat we are ascribing to it some character; and though it is natural to think that this character consists in a relation to a proposition; yet, since this view has been disputed, we shall start our inquiry from what is undoubtedly real, which is not the proposition but the character of propositional reference. We shall have to discuss its analysis later, but for our immediate purposes we can take it without analysis as something with which we are all familiar.

Propositional reference is not, of course, confined to beliefs; my knowledge that the earth is round, my opinion that free trade is superior to protection, any form of thinking, knowing, or being under the impression that—has a propositional reference, and it is only such states of mind that can be either true or false. Merely thinking of Napoleon cannot be true or false, unless it is thinking *that* he was or did so and so; for if the reference is not propositional, if it is not the sort of reference which it takes a sentence to express, there can be neither truth nor falsity. On the other hand not all states which have propositional reference are either true or false; I can hope it will be fine to-morrow, wonder whether it will be fine to-morrow, and finally believe it will be fine to-morrow. These three states all have the same propositional reference but only the belief can be called true or false. We do not call wishes, desires or wonderings true, not because they have no propositional reference, but because they lack what may be called an affirmative or assertive character, the element that is present in thinking that, but absent in wondering whether. In the absence of some degree of this character we never use the words true or false, though the degree need be only of the slightest and we can speak of an assumption as true, even when it is only made in order to discover its consequences. For states with the opposite character of denial we do not

naturally use the words true or false, though we can call them correct or incorrect according as beliefs with the same propositional reference would be false or true.

The mental states, [then], with which we are concerned, those, namely, with propositional reference and some degree of the affirmative character, have unfortunately no common name in ordinary language. There is no term applicable to the whole range from mere conjecture to certain knowledge, and I propose to meet this deficiency<sup>4</sup> by using the terms *belief* and *judgment* as synonyms to cover the whole range of [mental] states in question [although this involves a great widening of their ordinary meanings] and not in their ordinary narrower meanings.

It is, then, in regard to beliefs or judgments that we ask for the meaning of truth and falsity, and it seems advisable to begin by explaining that these are not just vague terms indicating praise or blame of any kind, but have a quite definite meaning. There are various respects in which a belief can be regarded as good or bad; it can be true or false, it can be held with a higher or a low degree of confidence, for good or for bad reasons, in isolation or as part of a coherent system of thought, and for any clear discussion to be possible it is essential to keep those forms of merit distinct from one another, and not to confuse them by using the word "true" in a vague way first for one and then for another. This is a point on which ordinary speech is sounder than the philosophers; to take an example of Mr Russell's, someone who thought that the present Prime Minister's name began with B would think so truly, even if he had derived his opinion from the mistaken idea that the Prime Minister was Lord Birkenhead; and it is clear that by calling a belief true, we neither mean nor imply that it is either well-grounded or comprehensive and that if these qualities are confused with truth as they are, for instance, by Bosanquet,<sup>5</sup> any profitable discussion of the subject becomes impossible. The kind of merit in a belief to which we refer in calling it true can be easily seen to be something which depends only on its propositional reference;<sup>6</sup> if one man's belief that the earth is round is true so is anyone else's belief that the earth is round, however little reason he may have for thinking so.

After these preliminaries we must come to the point: what *is* the meaning of 'true'? It seems to me that the answer is really perfectly obvious, that anyone can *see* what it is and that difficulty only arise when we try

to say what it is, because it is something which ordinary language is rather ill-adapted to express.

Suppose a man believes that the earth is round; then his belief is true because the earth *is* round; or generalising this, if he believes that *A* is *B* his belief will be true if *A* is *B* and false otherwise.

It is, I think, clear that in this last sentence we have the meaning of truth explained, and that the only difficulty is to formulate this explanation strictly as a definition. If we try to do this, the obstacle we encounter is that we cannot describe all beliefs as beliefs that *A* is *B* since the propositional reference of a belief may have any number of different more complicated forms. A man may be believing that all *A* are not *B*, or that if all *A* are *B*, either all *C* are *D* or some *E* are *F*, or something still more complicated. We cannot, in fact, assign any limit to the number of forms which may occur, and must therefore be comprehended in a definition of truth; so that if we try to make a definition to cover them all it will have to go on forever, since we must say that a belief is true, if supposing it to be a belief that *A* is *B*, *A* is *B*, or if supposing it to be a belief that *A* is not *B*, *A* is not *B*, or if supposing it to be a belief that either *A* is *B* or *C* is *D*, either *A* is *B* or *C* is *D*, and so on ad infinitum.

In order to avoid this infinity we must consider the general form of a propositional reference of which all these forms are species; any belief whatever we may symbolise as a belief that *p*, where '*p*' is a variable sentence just as '*A*' and '*B*' are variable words or phrases (or terms as they are called in logic). We can then say that a belief is true if it is a belief that *p*, and *p*.<sup>7</sup> This definition sounds odd because we do not at first realize that '*p*' is a variable *sentence* and so should be regarded as containing a verb; "and *p*" sounds nonsense because it seems to have no verb and we are apt to supply a verb such as "is true" which would of course make nonsense of our definition by apparently reintroducing what was to be defined. But '*p*' really contains a verb; for instance, it might be "*A* is *B*" and in this case we should end up "and *A* is *B*" which can as a matter of ordinary grammar stand perfectly well by itself.

The same point exactly arises if we take, not the symbol '*p*', but the relative pronoun which replaces it in ordinary language. Take for example "what he believed was true." Here what he believed was, of course, something expressed by a sentence containing a verb. But when we rep-



resent it by the pronoun 'what' the verb which is really contained in the 'what' has, as a matter of language, to be supplied again by "was true." If however we particularize the form of belief in question all need for the words "was true" disappears as before and we can say "the things he believed to be connected by a certain relation were, in fact,<sup>8</sup> connected by that relation."

As we claim to have defined truth we ought to be able to substitute our definition for the word 'true' wherever it occurs. But the difficulty we have mentioned renders this impossible in ordinary language which treats what should really be called *pro-sentences* as if they were *pro-nouns*. The only pro-sentences admitted by ordinary language are 'yes' and 'no', which are regarded as by themselves expressing a complete sense, whereas 'that' and 'what' even when functioning as short for sentences always require to be supplied with a verb: this verb is often "is true" and this peculiarity of language gives rise to artificial problems as to the nature of truth, which disappear at once when they are expressed in logical symbolism, in which we can render "what he believed is true" by "if  $p$  was what he believed,  $p$ ".

So far we have dealt only with truth; what about falsity? The answer is again simply expressible in logical symbolism, but difficult to explain in ordinary language. There is not only the same difficulty that there is with truth but an additional difficulty due to the absence in ordinary language of any simple uniform expression for negation. In logical symbolism, for any proposition symbol  $p$  (corresponding to a sentence), we form the contradictory  $\neg p$  (or  $\sim p$  in *Principia Mathematica*); but in English we often have no similar way of reversing the sense of a sentence without considerable circumlocution. We cannot do it merely by putting in a "not" except in the simplest cases; thus "The King of France is not clever" is ambiguous, but on its most natural interpretation means "There is a King of France but he is not clever" and so is not what we get by simply denying "The King of France is clever"; and in more complicated sentences such as "if he comes, she will come with him" we can only deny either by a method special to the particular form of proposition, like "if he comes, she will not *necessarily* come with him" or by the general method of prefixing "It is not true that \_\_\_\_," "it is false that \_\_\_\_" or "It is not the case that \_\_\_\_," where [again] it looks as if two new ideas,

'truth' and 'falsity', were involved, but in reality we are simply adopting a round-about way of applying *not* to the sentence as a whole.

Consequently our definition of falsity (to believe falsely is to believe  $p$ , when  $\neg p$ ) is doubly difficult to put into words; but to argue that it is circular, because it defines falsity in terms of the operation of negation which cannot always be rendered in language without using the word "false", would simply be a confusion. "False" is used in ordinary language in two ways: first as part of a way of expressing negation, correlative to the use of "true" as a purely stylistic addition (as when "it is true that the earth is round" means no more than that the earth is round); and secondly as equivalent to not true, applied to beliefs or other states of mind having propositional references or derivatively to sentences or other symbols expressing those states of mind. The use we are trying to define is the second, not the first, which in the guise of the symbol  $\neg p$  we are taking for granted and propose to discuss later under the head of negation.<sup>9</sup>

Our definition that a belief is true if it is a "belief that  $p$ " and  $p$ , but false if it is a "belief that  $p$ " and  $\neg p$  is, it may be remarked, substantially that of Aristotle, who considering only the two forms " $A$  is" and " $A$  is not" declared that "To say of what is, that it is not, or of what is not, that it is, is false, while to say of what is that it is, and of what is not that it is not, is true."<sup>10</sup>

Although we have not yet used the word 'correspondence' ours will probably be called a Correspondence Theory of Truth. For if  $A$  is  $B$  we can speak according to common usage of the fact that  $A$  is  $B$  and say that it corresponds to the belief that  $A$  is  $B$  in a way in which if  $A$  is not  $B$  there is no such fact corresponding to it. But we cannot describe the nature of this correspondence until we know the analysis of propositional reference, of "believing that  $A$  is  $B$ ." Only when we know the structure of belief can we say what type of correspondence it is that unites true beliefs and facts. And we may well be sceptical as to there being any simple relation of correspondence applicable to all cases or even if it is always right to describe the relation as holding between the "belief that  $p$ " and the "fact that  $p$ "; for instance if the belief is disjunctive as it is when Jones thinks that Smith is either a liar or a fool, are we to say that it is made true by a "disjunctive fact," "the fact," namely, "that Smith is

either a liar or a fool"? [If we believe that reality contains no such mere "either-or" we shall have to modify our account.] Or if we hold it absurd to believe that reality contains such a mere either-or, what does the belief correspond to?

But the prospect of these difficulties need not distress us or lead us to suppose that we are on a wrong track in adopting what is, in a vague sense, a correspondence theory of truth. For we have given a clear definition of truth which escapes all these difficulties by not appealing to a notion of correspondence at all. A belief that  $p$ , we say, is true if and only if  $p$ ; for instance a belief that Smith is either a liar or a fool is true if Smith is either a liar or a fool and not otherwise. It seems, indeed, possible to replace this definition by a periphrasis about the correspondence of two facts; but if such a periphrasis is not ultimately legitimate that does not prove that our definition is wrong, but merely that it should not strictly be called a correspondence theory and that a statement of it in terms of correspondence should be regarded as merely an inaccurate popular explanation. Truth, we say, is when a man believes that  $A$  is  $B$  and  $A$  is  $B$ , whether or not such an occurrence can be accurately described as a correspondence between two facts; failure to describe it in terms of correspondence cannot show that it never occurs and is not what we mean by truth.

This account of truth is merely a truism, but there is no platitude so obvious that eminent philosophers have not denied it, and at the risk of wearying the reader we shall insist on our truism once more.

Let us take three statements like this:

The earth is round.

It is true that the earth is round.

Anyone who believes that the earth is round believes truly.

It is really obvious that these statements are all equivalent, in the sense that it is not possible to affirm one of them and deny another without patent contradiction; to say, for instance, that it is true that the earth is round but that the earth is not round is plainly absurd.

Now the first statement of the three does not involve the idea of truth in any way, it says simply that the earth is round. [In the second we have to prefix "It is true that" which is generally added not to alter the meaning

but for what in a wide sense are reasons of style [and does not affect the meaning of the statements]. Thus we can use it rather like “although” in conceding a point but denying a supposed consequence, “It is true that the earth is round, but still...,” or again we often use it when what we say has been questioned: “Is that true?” “Yes, it is perfectly true.” [But in the last case the phrase “it is true that the earth is round” is changing from simply meaning that the earth is round ...]

The meaning of the second, on the other hand, is less clear: it may be a mere synonym for the first, but more often contains some reference to the possibility of someone believing or saying that the earth is round. We are thinking not merely that the earth is round, but that because it is round anyone<sup>11</sup> who believes or says that it is round believes or says truly. We have passed from the first of our statements to the third. But the third amounts in a sense to no more than the first, and it is merely the first thought of in connection with the possibility of someone saying or believing it. To take a parallel case, we can say simply “The weather in Scotland was bad in July”, or we can think of that fact in reference to its possible effect on one of our friends and say instead “If you were in Scotland in July, you had bad weather.” So too we can think of the earth being round as a possible subject of belief and say “If you think the earth is round, you think truly” and this amounts to no more than that the earth has the quality you think it has when you think it is round, i.e. that the earth is round.

All this is really so obvious that one is ashamed to insist on it, but our insistence is rendered necessary by the extraordinary way in which philosophers produce definitions of truth in no way compatible with our platitudes, definitions according to which the earth can be round without its being true that it is round.<sup>12</sup> The reason for this lies in a number of confusions of which it must be extremely hard to keep clear if we are to judge by their extraordinary prevalence. In the rest of this chapter we shall be occupied solely with the defence of our platitude that a belief that *p* is true if and only if *p*, and in an attempt to unravel the confusions that surround it.

The first type of confusion arises from the ambiguity of the question which we are trying to answer, the question “what is truth?”, which can be interpreted in at least three different ways. For in the first place there

are some philosophers who do not see any problem in what is *meant* by 'truth', but take our interpretation of the term as being obviously right, and proceed under the title of "what is truth?" to discuss the different problem of giving a general criterion for distinguishing truth and falsehood. This was for instance Kant's interpretation<sup>13</sup> and he goes on quite rightly to say that the idea of such a general criterion of truth is absurd, and that for men to discuss such a question is as foolish as for one to milk a he-goat while another holds a sieve to catch the milk.

And secondly even when we agree that the problem is to define truth in the sense of explaining its meaning, this problem can wear two quite different complexions according to the kind of definition with which we are prepared to be content. Our definition is one in terms of propositional reference, which we take as a term already understood. But it may be held that this notion of propositional reference is itself in need of analysis and definition, and that a definition of truth in terms of so obscure a notion represents very little if any progress. If a belief is identified as what Mr Jones was thinking at 10 o'clock in the morning, and we ask what is meant by calling the belief so identified a true belief, to apply the only answer we have so far obtained we need to know what Mr Jones' belief was a "belief that"; for instance, we say that if it was a belief that the earth is flat, then it was true if the earth is flat. But to many this may seem merely to shirk the hardest and most interesting part of the problem, which is to find out how and in what sense those images or ideas in Mr Jones' mind at 10 o'clock constitute or express a "belief that the earth is flat." Truth, it will be said, consists in a relation between ideas and reality, and the use without analysis of the term propositional reference simply conceals and shirks all the real problems that this relation involves.

This charge must be admitted to be just, and an account of truth which accepts the notion of propositional reference without analysis cannot possibly be regarded as complete. For all the many difficulties connected with that notion are really involved in truth which depends on it: if, for instance, "propositional reference" has quite different meanings in relation to different kinds of belief (as many people think) then a similar ambiguity is latent in 'truth' also, and it is obvious that we shall not have

got our idea of truth really clear until this and all similar problems are settled.

But though the reduction of truth to propositional reference is a very small part and much the easiest part of its analysis, it is not therefore one which we can afford to neglect. [Not only is it essential to realize that truth and propositional reference are not independent notions requiring separate analysis, and that it is truth that depends on and must be defined *via* reference not reference *via* truth.<sup>14</sup>] For not only is it in any event essential to realise that the problem falls in this way into two parts,<sup>15</sup> the reduction of truth to reference and the analysis of reference itself, and to be clear which part of the problem is at any time being tackled, but for many purposes it is only the first and easiest part of the solution that we required; we are often concerned not with beliefs or judgments as occurrences at particular times in particular men's minds, but with, for instance, *the* belief or judgment "all men are mortal"; in such case the only definition of truth we can possibly need is one in terms of propositional reference, which is presupposed in the very notion of *the* judgment "all men are mortal"; for when we speak of *the* judgment "all men are mortal" what we are really dealing with is any particular judgement on any particular occasion which has that propositional reference, which is a judgment "that all men are mortal." Thus, though the psychological difficulties involved in this notion of reference must be faced in any complete treatment of truth, it is well to begin with a definition which is sufficient for a great many purposes and depends only on the simplest considerations.

And whatever the complete definition may be, it must preserve the evident connection between truth and reference, that a belief "that *p*" is true if and only if *p*. We may deride this as trivial formalism, but since we cannot contradict it without absurdity, it provides a slight check on any deeper investigations that they must square with this obvious truism.

## Notes

1. How difficult the problem is may be judged from the fact that in the years 1904–25 Mr Bertrand Russell has adopted in succession five different solutions of it.
2. I use "state" as the widest possible term, not wishing to express any opinion as to the nature of beliefs etc.

3. Or, of course, that something is not so and so, or that if something is so and so, something is not such and such, and so on through all the possible forms.

4. [It should perhaps be remarked that the late Professor Cook Wilson held that these mental states do not in fact belong. (Square brackets indicate remarks F. P. Ramsey made on the manuscript—Ed.)] It should, however, be remarked that according to one theory this is not really a deficiency at all, since the states in question have nothing important in common. Knowledge and opinion have propositional reference in quite different senses and are not species of a common genus. This view, put forward most clearly by J. Cook Wilson, (but also implied by others, e.g., Edmund Husserl) is explained and considered below.

5. Bernard Bosanquet, *Logic*, 2nd ed., Vol II (Oxford, 1911), pp. 282 ff. Of course he sees the distinction but he deliberately blurs it, arguing that an account of truth which enables an ill-grounded statement to be true, cannot be right. His example of the man who makes a true statement believing it to be false, reveals an even grosser confusion. He asks why such a statement is a lie, and answers this by saying that “it was contrary to the system of his knowledge as determined by his whole experience at the time.” Granting this, it would at most follow that coherence with the man’s system of his knowledge is a mark not of *truth* (for *ex hypothesi* such a statement would have been false) but of *good faith*; and this is brought in as an argument in favour of a coherence theory of *truth*!

6. It has been suggested by Professor Moore (“Facts and Propositions,” *Proceedings of the Aristotelian Society*, Supplementary Volume VII (1927), pp. 171–206; see p. 178) that the same entity may be both a belief that (say) the earth is round and a belief that something else; in this case it will have two propositional references and may be true in respect of one and false in respect of the other. It is not to my mind a real possibility, but everything in the present chapter could easily be altered so as to allow for it, though the complication of language which would result seems to me far to outweigh the possible gain in accuracy.

7. In Mr Russell’s symbolism

$B$  is true  $\text{:=} (\exists p)$ .  $B$  is a belief that  $p \ \& \ p$ . Df

8. In a sentence like this “in fact” serves simply to show that the *oratio obliqua* introduced by “he believed” has now come to an end. It does not mean a new notion to be analyzed, but is simple a connecting particle.

9. See below. [Presumably this is a reference to the unwritten chapter on negation.—Ed.]

10. *Metaphysics*, Gamma, 6 1011b25, Mr Ross’ translation.

11. For instance the man we are talking to may have just made the point and we concede it. “Yes, it’s true, as you say, that the earth is round, but \_\_\_\_” or we may have made it and be questioned “Is that true, what you were saying, that the earth is round?” “Yes, it’s perfectly true.”

12. Thus according to William James a pragmatist could think both that Shakespeare’s plays were written by Bacon and that someone else’s opinion that

Shakespeare wrote them might be perfectly true “for him.” (“The Meaning of Truth,” p. 274.) On the idea that what is true for one person may not be true for another see below.

13. See *Kritik der reinen Vernunft*, “Die transzendente Logik.” Einleitung III (A57 = B82): “Die alte und berühmte Frage ... *Was ist Wahrheit?* Die Namen-erklärung der Wahrheit, dass sie nämlich die Übereinstimmung der Erkenntnis mit ihrem Gegenstande sei, wird hier geschenkt und vorausgesetzt; man verlangt aber zu wissen, welches das allgemeine und sichere Kriterium der Wahrheit einer jedem Erkenntnis sei.” The reason why there can be no such criterion is that every object is distinguishable and therefore has something true of it which is true of no other object. Hence there can be no guarantee of truth irrespective of the object in question.

14. [This might perhaps be denied if reference were something essentially different in the cases of true and of false beliefs; e.g., if the precise way in which a man’s belief today that it will be wet tomorrow was a belief “that it will be wet to-morrow” depended on how to-morrow’s weather actually turned out. But this is absurd for it would allow us to settle the weather in advance by simply considering the nature of the prophet’s expectation and seeing whether it had true-reference or false-reference.]

15. It might possibly be questioned whether this division of the problem is sound, not because the truth of a belief does not obviously depend on its reference, i.e., on what is believed, but because reference might be essentially different in the two cases of truth and falsity, so that there were really two primitive ideas, true-reference and false-reference, which had to be separately analysed. In this case, however, we could tell whether a belief that *A* is *B* were true or false, without looking at *A* by simply seeing whether the manner in which the belief was a “belief that *A* is *B*” was that of true-reference or false-reference, and infer with certainty that to-morrow would be fine from the fact that someone believed in a particular way, the way of false-reference, that it would be wet.





P. F. Strawson

Mr. Austin offers us a purified version of the correspondence theory of truth. On the one hand he disclaims the semanticists' error of supposing that "true" is a predicate of sentences; on the other, the error of supposing that the relation of correspondence is other than purely conventional, the error which models the word on the world or the world on the word. His own theory is, roughly, that to say that a statement is true is to say that a certain speech-episode is related in a certain conventional way to something in the world exclusive of itself. But neither Mr. Austin's account of the two terms of the truth-conferring relation, nor his account of the relation itself, seems to me satisfactory. The correspondence theory requires, not purification, but elimination.

## 1 Statements

It is, of course, indisputable that we use various substantival expressions as grammatical subjects of "true." These are, commonly, noun-phrases like "What he said" or "His statement"; or pronouns or nounphrases, with a "that"-clause in apposition, *e.g.*, "It . . . that *p*" and "The statement that *p*." Austin proposes that we should use "statement" to do general duty for such expressions as these. I have no objection. This will enable us to say, in a philosophically non-committal way, that, in using "true," we are talking about statements. By "saying this in a non-committal way," I mean saying it in a way which does not commit us to any view about the nature of statements so talked about; which does not commit us, for example, to the view that statements so talked about are historic events.

The words "assertion" and "statement" have a parallel and convenient duplicity of sense. "My statement" may be either what I say or my saying it. My saying something is certainly an episode. What I say is not. It is the latter, not the former, we declare to be true. (Speaking the truth is not a manner of speaking: it is saying something true.) When we say "His statement was received with thunderous applause" or "His vehement assertion was followed by a startled silence," we are certainly referring to, characterising, a historic event, and placing it in the context of others. If I say that the same statement was first whispered by John and then bellowed by Peter, uttered first in French and repeated in English, I am plainly still making historical remarks about utterance-occasions; but the word "statement" has detached itself from reference to any particular speech-episode. The episodes I am talking about are the whisperings, bellowings, utterings and repetitions. The statement is not something that figures in all these episodes. Nor, when I say that the statement is true, as opposed to saying that it was, in these various ways, made, am I talking indirectly about these episodes or any episodes at all. (Saying of a statement that it is true is not related to saying of a speech-episode that it was true as saying of a statement that it was whispered is related to saying of a speech-episode that it was a whisper.) It is futile to ask what thing or event I *am* talking about (over and above the subjectmatter of the statement) in declaring a statement to be true; for there is no such thing or event. The word "statement" and the phrase "What he said," like the conjunction "that" followed by a noun clause, are convenient, grammatically substantival, devices, which we employ, on certain occasions, for certain purposes, notably (but not only) the occasions on which we use the word "true." What these occasions are I shall try later to elucidate. To suppose that, whenever we use a singular substantive, we are, or ought to be, using it to refer to something, is an ancient, but no longer a respectable, error.

More plausible than the thesis that in declaring a statement to be true I am talking about a speech-episode is the thesis that in order for me to declare a statement true, there must have occurred, within my knowledge, at least one episode which was a making of that statement. This is largely, but (as Austin sees) not entirely, correct. The occasion of my declaring a statement to be true may be not that someone has made the

statement, but that I am envisaging the possibility of someone's making it. For instance, in discussing the merits of the Welfare State, I might say: "It is true that the general health of the community has improved (that *p*), but this is due only to the advance in medical science." It is not necessary that anyone should have said that *p*, in order for this to be a perfectly proper observation. In making it, I am not talking *about* an actual or possible speech-episode. I am myself asserting that *p*, in a certain way, with a certain purpose. I am anticipatorily conceding, in order to neutralize, a possible objection. I forestall someone's making the statement that *p* by making it myself, with additions. It is of prime importance to distinguish the fact that the use of "true" always glances backwards or forwards to the actual or envisaged making of a statement by someone, from the theory that it is used to characterise such (actual or possible) episodes.

It is not easy to explain the non-episodic and noncommittal sense of "statement" in which "statement" = "what is said to be true or false." But, at the risk of being tedious, I shall pursue the subject. For if Austin is right in the suggestion that it is basically of speech-episodes that we predicate "true," it should be possible to "reduce" assertions in which we say of a statement in the non-episodic sense that it is true to assertions in which we are predicating truth of episodes. Austin points out that the same sentence may be used to make different statements. He would no doubt agree that different sentences may be used to make the same statement. I am not thinking only of different languages or synonymous expressions in the same language; but also of such occasions as that on which you say of Jones "He is ill," I say *to* Jones "You are ill" and Jones says "I am ill." Using, not only different sentences, but sentences with different meanings, we all make "the same statement"; and this is the sense of "statement" we need to discuss, since it is, *prima facie*, of statements in this sense that we say that they are true or false (*e.g.*, "What they all said, namely, that Jones was ill, was quite true."). We could say: people make the same statement when the words they use in the situations in which they use them are such that they must (logically) either all be making a true statement or all be making a false statement. But this is to use "true" in the elucidation of "same statement." Or we could say, of the present case: Jones, you and I all make the same statement because,

using the words we used in the situation in which we used them, we were all applying the same description to the same person at a certain moment in his history; anyone applying that description to that person (etc.), would be making that statement. Mr. Austin might then wish to analyse (A) "The statement that Jones was ill is true" in some such way as the following: "If anyone has uttered, or were to utter, words such that in the situation in which they are uttered, he is applying to a person the same description as I apply to that person when I now utter the words 'Jones was ill,' then the resulting speech-episode was, or would be, true." It seems plain, however, that nothing but the desire to find a metaphysically irreproachable first term for the correspondence relation could induce anyone to accept this analysis of (A) as an elaborate general hypothetical. It would be a plausible suggestion only if the grammatical subjects of "true" were *commonly* expressions referring to particular, uniquely dateable, speech-episodes. But the simple and obvious fact is that the expressions occurring as such grammatical subjects ("What they said," "It ... that *p*" and so on) never do, in these contexts, stand for such episodes.<sup>1</sup> *What they said* has no date, though their several sayings of it are dateable. *The statement that p* is not an event, though it had to be made for the first time and made within my knowledge if I am to talk of its truth or falsity. If I endorse Plato's view, wrongly attributing it to Lord Russell ("Russell's view that *p* is quite true"), and am corrected, I have not discovered that I was talking of an event separated by centuries from the one I imagined I was talking of. (Corrected, I may say: "Well it's true, whoever said it.") My *implied* historical judgment is false; that is all.

## 2 Facts

What of the second term of the correspondence relation? For this Mr. Austin uses the following words or phrases: "thing," "event," "situation," "state of affairs," "feature" and "fact." All these are words which should be handled with care. I think that through failing to discriminate sufficiently between them, Mr. Austin (1) encourages the assimilation of facts to things, or (what is approximately the same thing) of stating to referring; (2) misrepresents the use of "true"; and (3) obscures another and more fundamental problem.

In section 3 of his paper, Mr. Austin says, or suggests, that all stating involves both referring (“demonstration”) and characterizing (“description”). It is questionable whether all statements do involve both,<sup>2</sup> though it is certain that some do. The following sentences, for example, could all be used to make such statements; *i.e.*, statements in the making of which both the referring and describing functions are performed, the performance of the two functions being approximately (though not exclusively) assignable to different parts of the sentences as uttered:—

The cat has the mange.

That parrot talks a lot.

Her escort was a man of medium build, clean-shaven, well-dressed and with a North Country accent.

In using such sentences to make statements, we refer to a thing or person (object) in order to go on to characterize it: (we demonstrate in order to describe). A *reference* can be correct or incorrect. A *description* can fit, or fail to fit, the thing or person to which it is applied.<sup>3</sup> When we refer correctly, there certainly is a conventionally established relation between the words, so used, and the thing to which we refer. When we describe-correctly, there certainly is a conventionally established relation between the words we use in describing and the type of thing or person we describe. These relations, as Mr. Austin emphasizes, are different. An expression used referringly has a different logical role from an expression used descriptively. They are differently related to the object. And *stating* is different from referring, and different from describing; for it is (in such cases) both these at once. Statement (*some* statement) is reference-cum-description. To avoid cumbersome phrasing, I shall speak henceforward of *parts* of statements (the referring part and the describing part); though parts of statements are no more to be equated with parts of sentences (or parts of speech-episodes) than statements are to be equated with sentences (or speech-episodes).

That (person, thing, etc.) to which the referring part of the statement refers, and which the describing part of the statement fits or fails to fit, is that which the statement is *about*. It is evident that there is nothing else in the world for the statement itself to be related to either in some further way of its own or in either of the different ways in which these different

parts of the statement are related to what the statement is about. And it is evident that the demand that there should be such a relatum is logically absurd: a logically fundamental type-mistake. But the demand for something in the world *which makes the statement true* (Mr. Austin's phrase), or *to which the statement corresponds when it is true*, is just this demand. And the answering theory that to say that a statement is true is to say that a speech-episode is conventionally related in a certain way to such a relatum reproduces the type-error embodied in this demand. For while we certainly say that a statement corresponds to (fits, is borne out by, agrees with) the facts, as a variant on saying that it is true, we *never* say that a statement corresponds to the thing, person, etc., it is about. What "makes the statement" that the cat has mange "true," is not the cat, but the *condition* of the cat, *i.e.*, the fact that the cat has mange. The only plausible candidate for the position of what (in the world) makes the statement true is the fact it states; but the fact it states is not something in the world.<sup>4</sup> It is not an object; not even (as some have supposed) a complex object consisting of one or more particular elements (constituents, parts) and a universal element (constituent, part). I can (perhaps) hand you, or draw a circle round, or time with a stop-watch the things or incidents that are referred to when a statement is made. Statements are about such objects; but they state facts. Mr. Austin seems to ignore the complete difference of type between, *e.g.*, "fact" and "thing"; to talk as if "fact" were just a very general word (with, unfortunately, some misleading features) for "event," "thing," etc., instead of being (as it is) both wholly different from these, and yet the only possible candidate for the desired non-linguistic correlate of "statement." Roughly: the thing, person, etc. referred to is the material correlate of the referring part of the statement; the quality or property the referent is said to "possess" is the *pseudo*-material correlate of its describing part; and the fact to which the statement "corresponds" is the *pseudo*-material correlate of the statement as a whole.

These points are, of course, reflected in the behaviour of the word "fact" in ordinary language; behaviour which Mr. Austin notes, but by which he is insufficiently warned. "Fact," like "true," "states" and "statement" is wedded to "that"-clauses; and there is nothing unholy about this union. Facts are known, stated, learnt, forgotten, overlooked,

commented on, communicated or noticed. (Each of these verbs may be followed by a “that”-clause or a “the fact that”-clause.) Facts are what statements (when true) state; they are not what statements are about. They are not, like things or happenings on the face of the globe, witnessed or heard or seen, broken or overturned, interrupted or prolonged, kicked, destroyed, mended or noisy. Mr. Austin notes the expression “fact that,” warns us that it may tempt us to identify facts with true statements and explains its existence by saying that for certain purposes in ordinary life we neglect, or take as irrelevant, the distinction between saying something true and the thing or episode of which we are talking. It would indeed be wrong—but not for Mr. Austin’s reasons—to identify “fact” and “true statement”; for these expressions have different roles in our language, as can be seen by the experiment of trying to interchange them in context. Nevertheless their roles—or those of related expressions—overlap. There is no nuance, except of style, between “That’s true” and “That’s a fact”; nor between “Is it true that . . . ?” and “Is it a fact that . . . ?”<sup>5</sup> But Mr. Austin’s reasons for objecting to the identification seem mistaken, as does his explanation of the usage which (he says) tempts us to make it. Because he thinks of a statement as something in the world (a speech-episode) and a fact as something else in the world (what the statement either “corresponds to” or “is about”), he conceives the distinction as of overriding importance in philosophy, though (surprisingly) sometimes negligible for ordinary purposes. But I can conceive of no occasion on which I could possibly be held to be “neglecting or taking as irrelevant” the distinction between, say, my wife’s bearing me twins (at midnight) and my saying (ten minutes later) that my wife had borne me twins. On Mr. Austin’s thesis, however, my announcing “The fact is that my wife has borne me twins” would be just such an occasion.

Elsewhere in his paper, Mr. Austin expresses the fact that there is no theoretical limit to what could truly be said about things in the world, while there are very definite practical limits to what human beings actually can and do say about them, by the remark that statements “always fit the facts more or less loosely, in different ways for different purposes.” But what could fit more perfectly the fact that it is raining than the statement that it is raining? Of course, statements and facts fit. They were made for each other. If you prise the statements off the world you prise



the facts off it too; but the world would be none the poorer. (You don't also prise off the world what the statements are about—for this you would need a different kind of lever.)

A symptom of Mr. Austin's uneasiness about facts is his preference for the expressions "situation" and "state of affairs"; expressions of which the character and function are a little less transparent than those of "fact." They are more plausible candidates for inclusion in the world. For while it is true that situations and states of affairs are not seen or heard (any more than facts are), but are rather *summed up* or *taken in at a glance* (phrases which stress the connection with statement and "that"-clause respectively), it is also true that there is a sense of "about" in which we do talk about, do describe, situations and states of affairs. We say, for example, "The international situation is serious" or "This state of affairs lasted from the death of the King till the dissolution of Parliament." In the same sense of "about," we talk about facts; as when we say "I am alarmed by the fact that kitchen expenditure has risen by 50 per cent in the last year." But whereas "fact" in such usages is linked with a "that"-clause (or connected no less obviously with "statement," as when we "take down the facts" or hand someone the facts on a sheet of paper), "situation" and "state of affairs" stand by themselves, states of affairs are said to have a beginning and an end, and so on. Nevertheless, situations and states of affairs so talked of are (like facts so talked of), abstractions that a logician, if not a grammarian, should be able to see through. Being alarmed by a fact is not like being frightened by a shadow. It is being alarmed because. . . . One of the most economical and pervasive devices of language is the use of substantival expressions to abbreviate, summarize and connect. Having made a series of descriptive statements, I can comprehensively connect with these the remainder of my discourse by the use of such expressions as "this situation" or "this state of affairs"; just as, having produced what I regard as a set of reasons for a certain conclusion I allow myself to draw breath by saying "Since *these things* are so, then . . .," instead of prefacing the entire story by the conjunction. A situation or state of affairs is, roughly, a set of facts not a set of things.

A point which it is important to notice in view of Mr. Austin's use of these expressions (in sections 3a and 3b of his paper) is that when we *do* "talk about" situations (as opposed to things and persons) the situation

we talk about is not, as he seems to think it is, correctly identified with the fact we state (with “what makes the statement true”). If a situation is the “subject” of our statement, then what “makes the statement true” is not the situation, but the fact that the situation has the character it is asserted to have. I think much of the persuasiveness of the phrase “talking about situations” derives from that use of the word on which I have just commented. But if a situation is treated as the “subject” of a statement, then it will not serve as the non-linguistic term, for which Mr. Austin is seeking, of the “relation of correspondence;” and if it is treated as the non-linguistic term of this relation, it will not serve as the subject of the statement.

Someone might now say: “No doubt ‘situation,’ ‘state of affairs,’ ‘facts’ are related in this way to ‘that’-clauses and assertive sentences; can serve, in certain ways and for certain purposes, as indefinite stand-ins for specific expressions of these various types. So also is ‘thing’ related to some nouns; ‘event’ to some verbs, nouns and sentences; ‘quality’ to some adjectives; ‘relation’ to some nouns, verbs and adjectives. Why manifest this prejudice in favour of things and events as alone being parts of the world or its history? Why not situations and facts as well?” The answer to this (implicit in what has gone before) is twofold.

(i)

The first part of the answer<sup>6</sup> is that the whole charm of talking of situations, states of affairs or facts as included in, or parts of, the world, consists in thinking of them as things, and groups of things; that the temptation to talk of situations, etc., in the idiom appropriate to talking of things and events is, once this first step is taken, overwhelming. Mr. Austin does not withstand it. He significantly slips in the word “feature” (noses and hills are *features*, of faces and landscapes) as a substitute for “facts.” He says that the reason why photographs and maps are not “true” in the way that statements are true is that the relation of a map or a photograph to what it is a map or a photograph of is not wholly (in the first case) and not at all (in the second) a conventional relation. But this is not the only, or the fundamental, reason (The relation between the Prime Minister of England and the phrase “the Prime Minister of England” is conventional; but it doesn’t make sense to say that someone uttering the phrase out of

context is saying something true or false.) The (for present purposes) fundamental reason is that “being a map of” or “being a photograph of” *are* relations, of which the non-photographic, non-cartographical, *relata* are, say, personal or geographical *entities*. The trouble with correspondence theories of truth is not primarily the tendency to substitute non-conventional relations for what is really a wholly conventional relation. It is the misrepresentation of “correspondence between statement and fact” *as a relation, of any kind, between events or things or groups of things* that is the trouble. Correspondence theorists think of a statement as “describing that which makes it true” (fact, situation, state of affairs) in the way a descriptive predicate may be used to describe, or a referring expression to refer to, a thing.<sup>7</sup>

(ii)

The second objection to Mr. Austin’s treatment of facts, situations, states of affairs as “parts of the world” which we declare to stand in a certain relation to a statement when we declare that statement true, goes deeper than the preceding one but is, in a sense, its point. Mr. Austin rightly says or implies (section 3) that for some of the purposes for which we use language, there must be conventions correlating the words of our language with what is to be found in the world. Not all the linguistic purposes for which this necessity holds, however, are identical. Orders, as well as information, are conventionally communicated. Suppose “orange” always meant what we mean by “Bring me an orange” and “that orange” always meant what we mean by “Bring me that orange,” and, in general, our language contained only sentences in some such way imperative. There would be no less need for a conventional correlation between the word and the world. Nor would there be any less to be found in the world. But those pseudo-entities which *make statements true* would not figure among the non-linguistic correlates. They would no more be found; (they never were found, and never did figure among the non-linguistic correlates). The point is that the word “fact” (and the “set-of-facts” words like “situation” “state of affairs”) have, like the words “statement” and “true” themselves, a certain type of word-world-relating discourse (the informative) *built in* to them. The occurrence in ordinary discourse of the words “fact” “statement” “true” signals the occurrence of this

type of discourse; just as the occurrence of the words “order” “obeyed” signalizes the occurrence of another kind of conventional communication (the imperative). If our task were to elucidate the nature of the first type of discourse, it would be futile to attempt to do it in terms of the words “fact,” “statement,” “true,” for these words contain the problem, not its solution. It would, for the same reason, be equally futile to attempt to elucidate any one of these words (in so far as the elucidation of *that* word would be the elucidation of *this* problem) in terms of the others. And it is, indeed, very strange that people have so often proceeded by saying “Well, we’re pretty clear what a statement is, aren’t we? Now let us settle *the further question, viz.*, what it is for a statement to be true.” This is like “Well, we’re clear about what a command is: now what is it for a command to be obeyed?” As if one could divorce statements and commands from the point of making or giving them!

Suppose we had in our language the word “execution” meaning “action which is the carrying out of a command.” And suppose someone asked the philosophical question: What is *obedience*? What is it for a command to be *obeyed*? A philosopher might produce the answer: “Obedience is a conventional relation between a command and an execution. A command is obeyed when it corresponds to an execution.”

This is the Correspondence Theory of Obedience. It has, perhaps, a little less value as an attempt to elucidate the nature of one type of communication than the Correspondence Theory of Truth has as an attempt to elucidate that of another. In both cases, the words occurring in the solution incorporate the problem. And, of course, this intimate relation between “statement” and “fact” (which is understood when it is seen that they both incorporate this problem) explains why it is that when we seek to explain *truth* on the model of naming or classifying or any other kind of conventional or non-conventional relation between one thing and another, we always find ourselves landed with “fact,” “situation,” “state of affairs” as the non-linguistic terms of the relation.

But why should the problem of Truth (the problem about our use of “true”) be seen as this problem of elucidating the fact-stating type of discourse? The answer is that it shouldn’t be; but that the Correspondence Theory can only be fully seen through when it is seen as a barren attempt on this second problem. Of course, a philosopher concerned

with the second problem, concerned to elucidate a certain general type of discourse, must stand back from language and talk about the different ways in which utterances are related to the world (though he must get beyond “correspondence of statement and fact” if his talk is to be fruitful). But—to recur to something I said earlier—the occurrence *in ordinary discourse* of the words “true,” “fact,” etc., signalizes, without commenting on, the occurrence of a certain way of using language. When we use these words in ordinary life, we are talking within, and not about, a certain frame of discourse; we are precisely not talking about the way in which utterances are, or may be, conventionally related to the world. We are talking about persons and things, but in a way in which we could not talk about them if conditions of certain kinds were not fulfilled. The problem about the use of “true” is to see how this word fits into that frame of discourse. The surest route to the wrong answer is to confuse this problem with the question: What type of discourse is this?<sup>8</sup>

### 3 Conventional Correspondence

It will be clear from the previous paragraph what I think wrong with Mr. Austin’s account of the relation itself, as opposed to its terms. In section 4 of his paper he says that, when we declare a statement to be true, the relation between the statement and the world which our declaration “asserts to obtain” is “a purely conventional relation” and “one which we could alter at will.” This remark reveals the fundamental confusion of which Mr. Austin is guilty between:—

- a. the semantic conditions which must be satisfied for the statement that a certain statement is true to be itself true; and
- b. what is asserted when a certain statement is stated to be true.

Suppose A makes a statement, and B declares A’s statement to be true. Then for B’s statement to be true, it is, *of course*, necessary that the words used by A in making the statement should stand in a certain conventional (semantical) relationship with the world; and that the “linguistic rules” underlying this relationship should be rules “observed” by both A and B. It should be remarked that these conditions (with the exception of the condition about B’s observance of linguistic rules) are equally necessary conditions of A’s having made a true statement in using the words

he used. *It is no more and no less absurd to suggest that B, in making his statement, asserts that these semantic conditions are fulfilled than it is to suggest that A, in making his statement, asserts that these semantic conditions are fulfilled (i.e., that we can never use words without mentioning them).* If Mr. Austin is right in suggesting that to say that a statement is true is to say that “the historic state of affairs to which it [*i.e.*, for Mr. Austin, the episode of making it] is correlated by the demonstrative conventions (the one it ‘refers to’) is of a type with which the sentence used in making the statement is correlated by the descriptive conventions,” *then* (and this is shown quite clearly by his saying that the relation we assert to obtain is a “purely conventional one” which “could be altered at will”) in declaring a statement to be true, we are either:—

- a. talking about the meanings of the words used by the speaker whose making of the statement is the occasion for our use of “true” (*i.e.*, profiting by the occasion to give semantic rules); or
- b. saying that the speaker has used correctly the words he did use.

It is *patently* false that we are doing either of these things. Certainly, we use the word “true” when the semantic conditions described by Austin<sup>9</sup> are fulfilled; but we do not, in using the word, *state* that they are fulfilled. (And this, incidentally, is the answer to the question with which Mr. Austin concludes his paper.) The damage is done (the two problems distinguished at the end of the previous section confused) by asking the question: *When* do we use the word “true”? instead of the question: *How* do we use the word “true”?

Someone says: “It’s true that French Governments rarely last more than a few months, but the electoral system is responsible for that.” Is the fact he states in the first part of his sentence alterable by changing the conventions of language? It is not.

#### 4 Uses of “That”-Clauses; and of “Statement,” “True,” “Fact,” “Exaggerated,” Etc.

##### (a)

There are many ways of making an assertion about a thing, X, besides the bare use of the sentence-pattern “X is Y.” Many of these involve the use of “that”-clauses. For example:—

How often shall I have to tell you	}	that X is Y.
Today I learnt		
It is surprising		
The fact is		
I have just been reminded of the fact		
It is indisputable		
It is true		
It is established beyond question		

These are all ways of asserting, in very different context and circumstances, that X is Y.<sup>10</sup> Some of them involve autobiographical assertions as well; others do not. In the grammatical sense already conceded, all of them are “about” facts or statements. In no other sense is any of them about either, though some of them carry *implications* about the *making* of statements.

(b)

There are many different circumstances in which the simple sentence-pattern “X is Y” may be used to do things which are not merely stating (though they all involve stating) that X is Y. In uttering words of this simple pattern we may be encouraging, reproving or warning someone; reminding someone; answering, or replying to, someone; denying what someone has said; confirming, granting, corroborating, agreeing with, admitting what someone has said. Which of these, if any, we are doing depends on the circumstances in which, using this simple sentence-pattern, we assert that X is Y.

(c)

In many of the cases in which we are doing something besides merely stating that X is Y, we have available, for use in suitable contexts, certain abbreviatory devices which enable us to state that X is Y (to make our denial, answer, admission or whatnot) *without* using the sentence-pattern “X is Y.” Thus, if someone asks us “Is X Y?”, we may state (in the way of reply) that X is Y by saying “Yes.” If someone says “X is Y,” we may state (in the way of denial) that X is not Y, by saying “It is not” or by saying “That’s not true”; or we may state (in the way of corroboration, agreement, granting, etc.) that X is Y by saying “It is indeed” or

"That is true." In all these cases (of reply, denial and agreement) the context of our utterance, as well as the words we use, must be taken into account if it is to be clear what we are asserting, *viz.*, that X is (or is not) Y. It seems to me plain that in these cases "true" and "not true" (we rarely use "false") are functioning as abbreviatory statement-devices of the same general kind as the others quoted. And it seems also plain that the *only* difference between these devices which might tempt us, while saying of some ("Yes," "It is indeed," "It is not") that, in using them, we were talking about X, to say of others ("That's true," "That's not true") that, in using them, we were talking about something quite different, *viz.*, the utterance which was the occasion for our use of these devices, is their difference in grammatical structure, *i.e.*, the fact that "true" occurs as a grammatical predicate.<sup>11</sup> (It is obviously not a predicate of X.) If Mr. Austin's thesis, that in using the word "true" we make an assertion about a statement, were no more than the thesis that the word "true" occurs as a grammatical predicate, with, as grammatical subjects, such words and phrases as "That," "What he said," "His statement," etc., then, of course, it would be indisputable. It is plain, however, that he means more than this, and I have already produced my objections to the more that he means.

(d)

It will be clear that, in common with Mr. Austin, I reject the thesis that the phrase "is true" is logically superfluous, together with the thesis that so say that a proposition is true is *just* to assert it and to say that it is false is *just* to assert its contradictory. "True" and "not true" have jobs of their own to do, *some*, but by no means all, of which I have characterized above. In using them, we are not *just* asserting that X is Y or that X is not Y. We are asserting this in a way in which we could not assert it unless certain conditions were fulfilled; we may also be granting, denying, confirming, etc. It will be clear also that the rejection of these two theses does not entail acceptance of Mr. Austin's thesis that in using "true" we are making an assertion about a statement. Nor does it entail the rejection of the thesis which Mr. Austin (in Section 4 of his paper) couples with these two, *viz.*, the thesis that to say that an assertion is true



is not to make any further *assertion* at all. This thesis holds for many uses, but requires modification for others.

(e)

The occasions for using “true” mentioned so far in this section are evidently not the only occasions of its use. There is, for example, the generally concessive employment of “It is true that  $p$  . . .”, which it is difficult to see how Mr. Austin could accommodate. All these occasions have, however, a certain contextual immediacy which is obviously absent when we utter such sentences as “What John said yesterday is quite true” and “What La Rochefoucauld said about friendship is true.” Here the context of our utterance does not identify for us the statement we are talking about (in the philosophically non-committal sense in which we *are* “talking about statements” when we use the word “true”), and so we use a descriptive phrase to do the job. But the descriptive phrase does not identify an event; though the statement we make carries the implication (in some sense of “implication”) that there occurred an event which was John’s making yesterday (or Rochefoucauld’s making sometime) the statement that  $p$  (*i.e.*, the statement we declare to be true). We are certainly not telling our audience that the event occurred, *e.g.*, that John made the statement that  $p$ , for (i) we do not state, either by way of quotation or otherwise, what it was that John said yesterday, and (ii) our utterance achieves its main purpose (that of making, by way of confirmation or endorsement, the statement that  $p$ ) only if our audience already knows that John yesterday made the statement that  $p$ . The abbreviatory function of “true” in cases such as these becomes clearer if we compare them with what we say in the case where (i) we want to assert that  $p$ ; (ii) we want to indicate (or display our knowledge that) an event occurred which was John’s making yesterday the statement that  $p$ ; (iii) we believe our audience ignorant or forgetful of the fact that John said yesterday that  $p$ . We then use the formula “As John said yesterday,  $p$ ” or “It is true, as John said yesterday, that  $p$ ,” or “What John said yesterday, namely that  $p$ , is true.” (Of course the words represented by the letter  $p$ , which we use, may be—sometimes, if we are to make the same statement, must be—different from the words which John used.) Sometimes, to embarrass, or test, our audience, we use, in cases where the third of

these conditions is fulfilled, the formula appropriate to its nonfulfilment, *viz.*, "What John said yesterday is true."

(f)

In criticism of my view of truth put forward in *Analysis*,<sup>12</sup> and presumably in support of his own thesis that "true" is used to assert that a certain relation obtains between a speech-episode and something in the world exclusive of that episode, Mr. Austin makes, in Section 7 of his paper, the following point. He says: "Mr. Strawson seems to confine himself to the case when I say 'Your statement is true' or something similar—but what of the case when you state that S and I say nothing, but *look and see* that your statement is true?" The point of the objection is, I suppose, that since I *say* nothing, I cannot be making any performative use of "true"; yet I can see *that* your statement is true. The example, however, seems to have a force precisely contrary to what Mr. Austin intended. Of course, "true" has a different role in "X sees that Y's statement is true" from its role in "Y's statement is true." What is this role? Austin says in my hearing "There is a cat on the mat" and I look and see a cat on the mat. Someone (Z) reports: "Strawson saw that Austin's statement was true." What is he reporting? He is reporting that I have seen a cat on the mat; but he is reporting this in a way in which he could not report it except in certain circumstances, *viz.*, in the circumstances of Austin's having said in my hearing that there was a cat on the mat. Z's remark also carries the implication that Austin made a statement, but cannot be regarded as *reporting* this by implication since it fulfils its main purpose only if the audience already knows that Austin made a statement and what statement he made; and the implication (which *can* be regarded as an implied report) that I heard and understood what Austin said.<sup>13</sup> The man who looks and sees that the statement that there is a cat on the mat is true, sees no more and no less than the man who looks and sees that there is a cat on the mat, or the man who looks and sees that there is *indeed* a cat on the mat. But the *settings* of the first and third cases may be different from that of the second.

This example has value, however. It emphasizes the importance of the concept of the "occasion" on which we may make use of the assertive device which is the subject of this symposium (the word "true"); and

minimizes (what I was inclined to over-emphasize) the performatory character of our uses of it.

(g)

Mr. Austin stresses the differences between negation and falsity; rightly, in so far as to do so is to stress the difference (of occasion and context) between asserting that X is not Y and denying the assertion that X is Y. He also exaggerates the difference; for, if I have taken the point of his example, he suggests that there are cases in which "X is not Y" is inappropriate to a situation in which, if anyone stated that X was Y, it would be correct to say that the statement that X was Y was false. These are cases where the question of whether X is or is not Y does not arise (where the conditions of its arising are not fulfilled). They are equally, it seems to me, cases when the question of the truth or falsity of the statement that X is Y does not arise.

(h)

A qualification of my general thesis, that in using "true" and "untrue" we are not talking about a speech episode, is required to allow for those cases where our interest is not primarily in what the speaker asserts, but in the speaker's asserting it, in, say, the fact of his having *told the truth* rather than in the fact which he reported in doing so. (We may, of course, be interested in both; or our interest in a man's evident truthfulness on one occasion may be due to our concern with the degree of his reliability on others.)

But this case calls for no special analysis and presents no handle to any theorist of truth; for to use "true" in this way is simply to characterize a certain *event as the making*, by someone, of a true statement. The problem of analysis remains.

(i)

Mr. Austin says that we shall find it easier to be clear about "true" if we consider other adjectives "in the same class," such as "exaggerated," "vague," "rough," "misleading," "general," "too concise." I do not think these words *are* in quite the same class as "true" and "false." In any language in which statements can be made at all, it must be possible to

make true and false statements. But statements can suffer from the further defects Mr. Austin mentions only when language has attained a certain richness. Imagine one of Mr. Austin's rudimentary languages with "single words" for "complex situations" of totally different kinds. One could make true or false statements; but not statements which were exaggerated, over-concise, too general or rather rough. And even given a language as rich as you please, whereas all statements made in it could be true or false, not all statements could be exaggerated. When can we say that the statement that  $p$  is exaggerated? *One* of the conditions is this: that, if the sentence  $S_1$  is used to make the statement that  $p$ , there should be some sentence  $S_2$  (which could be used to make the statement that  $q$ ) such that  $S_1$  and  $S_2$  are related somewhat as "There were 200 people there" is related to "There were 100 people there." (To the remark "We got married yesterday," you cannot, except as a joke, reply: "You're exaggerating.")

Mr. Austin's belief, then, that the word "exaggerated" stands for a relation between a statement and something in the world exclusive of the statement, would at least be an over-simplification, even if it were not objectionable in other ways. But it is objectionable in other ways. The difficulties about statement and fact recur; and the difficulties about the relation. Mr. Austin would not want to say that the relation between an exaggerated statement and the world was like that between a glove and a hand too small for it. He would say that the relation was a conventional one. But the fact that the statement that  $p$  is exaggerated is not in any sense a conventional fact. (It is, perhaps, the fact that there were 1,200 people there and not 2,000.) If a man says: "There were at least 2,000 people there," you may reply (A) "No, there were not so many (far more)," or you may reply (B) "That's an exaggeration (understatement)." (A) and (B) say the same thing. Look at the situation more closely. In saying (A), you are not merely asserting that there were fewer than 2,000 people there: you are also correcting the first speaker, and correcting him in a certain general way, which you could not have done if he had not spoken as he did, though you could merely have asserted that there were fewer than 2,000 there without his having spoken. Notice also that what is being asserted by the use of (A)—that there were fewer than 2,000 there—cannot be understood without taking into account the

original remark which was the occasion for (A). (A) has both contextually-assertive and performatory features. (B) has the same features, and does the same job as (A), but more concisely and with greater contextual reliance.

Not all the words taken by Austin as likely to help us to be clear about “true” are in the same class as one another. “Exaggerated” is, of those he mentions, the one most relevant to his thesis; but has been seen to yield to my treatment. Being “over-concise” and “too general” are not ways of being “not quite true.” These obviously relate to the specific purposes of specific makings of statements; to the unsatisfied wishes of specific audiences. No alteration in things in the world, nor any magical replaying of the course of events, could bring statements so condemned into line, in the way that an “exaggerated assessment” of the height of a building could be brought into line by inorganic growth. Whether the statement (that *p*) is true or false is a matter of the way things are (of whether *p*); whether a statement is exaggerated (if the question arises—which depends on the type of statement and the possibilities of the language) is a matter of the way things are (*e.g.*, of whether or not there were fewer than 2,000 there). But whether a statement is over-concise<sup>14</sup> or too general depends on what the hearer wants to know. The world does not demand to be described with one degree of detail rather than another.

## 5 The Scope of “Statement,” “True,” “False” and “Fact”

Commands and questions, obviously do not claim to be statements of fact: they are not true or false. In Section 6 of his paper, Mr. Austin reminds us that there are many expressions neither interrogative nor imperative in form which we use for other purposes than that of reportage or forecast. From our employment of these expressions he recommends that we withhold (suspects that we do, in practice, largely withhold) the appellation “stating facts,” the words “true” and “false.” Philosophers, even in the sphere of language are not legislators; but I have no wish to challenge the restriction, in some philosophical contexts, of the words “statement,” “true,” “false,” to what I have myself earlier called the “fact-stating” type of discourse.

What troubles me more is Mr. Austin's own incipient analysis of this type of discourse. It seems to me such as to force him to carry the restriction further than he wishes or intends. And here there are two points which, though connected, need to be distinguished. First, there are difficulties besetting the relational theory of truth as such; second, there is the persistence of these difficulties in a different form when this "theory of truth" is revealed as, rather, an incipient analysis of the statement-making use of language.

First then, facts of the cat-on-the-mat-type are the favoured species for adherents of Mr. Austin's type of view. For here we have one thing (one chunk of reality) sitting on another: we can (if we are prepared to commit the errors commented on in Section (2) above) regard the two together as forming a single chunk, if we like, and call it a fact or state of affairs. The view may then seem relatively plausible that to say that the statement (made by me to you) that the cat is on the mat is true is to say that the three-dimensional state of affairs with which the episode of my making the statement is correlated by the demonstrative conventions is of a type with which the sentence I use is correlated by the descriptive conventions. Other species of fact, however, have long been known to present more difficulty: the fact that the cat is not on the mat, for example, or the fact that there are white cats, or that cats persecute mice, or that if you give my cat an egg, it will smash it and eat the contents. Consider the simplest of these cases, that involving negation. With what type of state-of-affairs (chunk of reality) is the sentence "The cat is not on the mat" correlated by conventions of description? With a mat *simpliciter*? With a dog on a mat? With a cat up a tree? The amendment of Mr. Austin's view to which one might be tempted for negative statements (*i.e.*, "S is true" = "The state of affairs to which S is correlated by the demonstrative conventions is *not* of a type with which *the affirmative form* of S is correlated by the descriptive conventions") destroys the simplicity of the story by creating the need for a different sense of "true" when we discuss negative statements. And worse is to follow. Not all statements employ conventions of demonstration. Existential statements don't, nor do statements of (even relatively) unrestricted generality. Are we to deny that these are statements, or create a further sense of "true"? And what has become

of the non-linguistic correlate, the chunk of reality? It this, in the case of existential or general statements, the entire world? Or, in the case of negatively existential statements, an ubiquitous non-presence?

As objections to a correspondence theory of truth, these are familiar points; though to advance them as such is to concede too much to the theory. What makes them of interest is their power to reveal how such a theory, in addition to its intrinsic defects, embodies too narrow a conception of the fact-stating use of language. Mr. Austin's description of the conditions under which a statement is true, regarded as an analysis of the fact-stating use, applies only to affirmative subject-predicate statements, *i.e.*, to statements in making which we refer to some one or more localized thing or group of things, event or set of events, and characterize it or them in some positive way (identify the object or objects and affix the label). It does not apply to negative, general and existential statements nor, straightforwardly, to hypothetical and disjunctive statements. I agree that any language capable of the fact-stating use must have some devices for performing the function to which Mr. Austin exclusively directs his attention, and that other types of statements of fact can be understood only in relation to this type. But the other types *are* other types. For example, the word "not" can usefully be regarded as a kind of crystallizing-out of something *implicit* in all use of descriptive language (since no predicate would have any descriptive force if it were compatible with everything). But from this it does not follow that negation (*i.e.*, the *explicit* exclusion of some characteristic) is a kind of affirmation, that negative statements are properly discussed in the language appropriate to affirmative statements. Or take the case of existential statements. Here one needs to distinguish two kinds of demonstration or reference. There is, first, the kind whereby we enable our hearer to identify the thing or person or event or set of these which we then go on to characterize in some way. There is, second, the kind by which we simply indicate a locality. The first ("Tabby has the mange") answers the question "Who which one, what) are you talking about?" The second ("There's a cat) the question "Where?" It is plain that no part of an existential statement performs the first function; though Austin's account of reference-cum-description is appropriate to reference of this kind rather than to that of the other. It is clear also that a good many existential statements do not

answer the question “Where?” though they may license the enquiry. The difference between various types of statement, and their mutual relations, is a matter for careful description. Nothing is gained by lumping them all together under a description appropriate only to one, even though it be the basic, type.

## 6 Conclusion

My central objection to Mr. Austin’s thesis is this. He describes the conditions which must obtain if we are correctly to declare a statement true. His detailed description of these conditions is, with reservations, correct as far as it goes, though in several respects too narrow. The central mistake is to suppose that in using the word “true” we are asserting such conditions to obtain. That this is a mistake is shown by the detailed examination of the behaviour of such words as “statement,” “fact,” etc., and of “true” itself, and by the examination of various different types of statement. This also reveals some of the ways in which “true” actually functions as an assertive device. What supremely confuses the issue is the failure to distinguish between the task of elucidating the nature of a certain type of communication (the empirically informative) from the problem of the actual functioning of the word “true” within the framework of that type of communication.

## Notes

Editor’s note: Strawson’s position on truth is further developed in the following works, where he places less emphasis on the confirmatory use of “true.”

“A Problem about Truth: A Reply to Mr. Warnock.” Originally appeared in *Truth*, ed. G. Pitcher. Englewood Cliffs: Prentice-Hall, 1964. Reprinted in *Logico-linguistic Papers* by P. F. Strawson.

“Reply to Searle.” In *The Philosophy of P. F. Strawson*, The Library of Living Philosophers, no. 26, ed. L. E. Hahn. Chicago: Open Court Publishing, 1998.

1. And the cases where such phrases might most plausibly be exhibited as having an episode-referring rôle are precisely those which yield most readily to another treatment; *viz.*, those in which one speaker corroborates, confirms or grants what another has just said (*see* Section 4 below).

2. *See* Section 5 below. The thesis that all statements involve both demonstration and description is, roughly, the thesis that all statements are, or involve, subject-predicate statements (not excluding relational statements).



3. Cf. the phrase "He is described as ..." What fills the gap is not a sentence (expression which could normally be used to make a statement), but a phrase which could occur as a part of an expression so used.

4. This is not, of course, to deny that there is that in the world which a statement of this kind is about (true or false *of*), which is *referred to* and *described* and which the description fits (if the statement is true) or fails to fit (if it is false). This truism is an inadequate introduction to the task of elucidating, not our use of "true," but a certain general way of using language, a certain type of discourse, *viz.*, the fact-stating type of discourse. What confuses the issue about the use of the word "true" is precisely its entanglement with this much more fundamental and difficult problem. (See (ii) of this section.)

5. I think in general the difference between them is that while the use of "true," as already acknowledged, glances backwards or forwards at an actual or envisaged making of a statement, the use of "fact" does not generally do this though it may do it sometimes. It certainly does not do it in, *e.g.*, the phrase "The fact is that ..." which serves rather to prepare us for the unexpected and unwelcome.

6. Which could be more shortly expressed by saying that if we read "world" (a sadly corrupted word) as "heavens and earth," talk of facts, situations and states of affairs, as "included in" or "parts of" the world is, obviously, metaphorical. The world is the totality of things, not of facts.

7. Suppose the pieces set on a chessboard, a game in progress. And suppose someone gives, in words, an exhaustive statement of the position of the pieces. Mr. Austin's objection (or one of his objections) to earlier correspondence theories is that they would represent the relation between the description and the board with the pieces on it as like, say, the relation between a newspaper diagram of a chess-problem and a board with the pieces correspondingly arranged. He says, rather, that the relation is a purely conventional one. My objection goes farther. It is that there is no thing or event called "a statement" (though there is the making of the statement) and there is no thing or event called "a fact" or "situation" (though there is the chessboard with the pieces on it) which stand to one another in any, even a purely conventional, relation as the newspaper diagram stands to the board-and-pieces. The facts (situation, state of affairs) cannot, like the chessboard-and-pieces, have coffee spilt on them or be upset by a careless hand. It is because Mr. Austin needs such events and things for his theory that he takes the making of the statement as the statement, and that which the statement is about as the fact which it states.

Events can be dated and things can be located. But the facts which statements (when true) state can be neither dated or located. (Nor can the statements, though the making of them can be.) Are they included in the world?

8. A parallel mistake would be to think that in our ordinary use (as opposed to a philosopher's use) of the word "quality," we were talking about people's uses of words; on the ground (correct in itself) that this word would have no use but for the occurrence of a certain general way of using words.

9. In what, owing to his use of the words “statement” “fact” “situation,” etc., is a misleading form. The quoted account of the conditions of truthful statement is more nearly appropriate as an account of the conditions of correct descriptive reference. Suppose, in a room with a bird in a cage, I say “That parrot is very talkative.” Then my use of the referring expression (“That parrot”) with which my sentence begins is correct when the token-object (bird) with which my token-expression (event) is correlated by the conventions of demonstration is of a kind with which the type-expression is correlated by the conventions of description. Here we do have an event and a thing and a (type-mediated) conventional relation between them. If someone corrects me, saying “That’s not a parrot; it’s a cockatoo,” he may be correcting either a linguistic or a factual error on my part. (The question of which he is doing is the question of whether I would have stuck to my story on a closer examination of the bird.) Only in the former case is he declaring a certain semantic condition to be unfulfilled. In the latter case, he is talking about the bird. He asserts that it is a cockatoo and not a parrot. This he could have done whether I had spoken or not. He also *corrects* me, which he could not have done if I had not spoken.

10. One might prefer to say that in some of these cases one was asserting only *by implication* that X is Y; though it seems to me more probable that in all these cases we should say, of the speaker, not “What he said implied that X is Y,” but “He *said* that X was Y.”

11. Compare also the English habit of making a statement followed by an interrogative appeal, for agreement in such forms as “isn’t it?”, “doesn’t he?” etc., with the corresponding German and Italian idioms, “Nicht wahr?”, “non è vero?” There is surely no significant difference between the phrases which do not employ the word for “true” and those which do: they all appeal for agreement in the same way.

12. Vol. 9, No. 6, June, 1949.

13. If I report: “I see that Austin’s statement is true,” this is simply a first-hand corroborative report that there is a cat on the mat, made in a way in which it could not be made except in these circumstances.

14. “Concise” is perhaps less often used of what a man says than of the way he says it (e.g., “concisely put,” “concisely expressed,” “a concise formulation”). A may take 500 words to say what B says in 200. Then I shall say that B’s formulation was more concise than A’s, meaning simply that he used fewer words.



---

## Truth

W. V. O. Quine

### 1 Vehicles of Truth

What are true or false, it will be widely agreed, are propositions. But it would not be so widely agreed were it not for ambiguity of ‘proposition’. Some understand the word as referring to sentences meeting certain specifications. Others understand it as referring rather to the meanings of such sentences. What looked like wide agreement thus resolves into two schools of thought: for the first school the vehicles of truth and falsity are the sentences, and for the second they are the meanings of the sentences.

A weakness of this second position is the tenuousness of the notion of sentence meanings. The tenuousness reaches the breaking point if one is persuaded of my thesis of the indeterminacy of translation (§§18, 21 [in *The Pursuit of Truth*—Ed.]). Even apart from that thesis, it seems perverse to bypass the visible or audible sentences and to center upon sentence meanings as truth vehicles; for it is only by recurring to the sentence that we can say which sentence meaning we have in mind.

There was indeed a motive for pressing to the sentence meanings. Many sentences in the same or different languages are deemed to be alike in meaning, and distinctions among them are indifferent to truth; so one narrowed the field by ascribing truth rather to the meanings. This motive would be excellent if the notion of sentence meaning were not so elusive. But as matters stand we fare better by treating directly of sentences. These we can get our teeth into.

There was also a second motive, equal and opposite to the first, for pressing on to the sentence meanings; namely, that one and the same sentence can be true on some occasions and false on others. Thus ‘The

Pope will visit Boston' was true but turned false after his last visit. 'I have a headache' is true or false depending on who says it and when. Ambiguity or vagueness of terms, also, can cause the truth value of a sentence to depend in part on the speaker's intention.

Propositions, thought of as sentence meanings, were the meanings exclusively of sentences of a firmer sort, not subject to such vacillations; what we may call *eternal* sentences.<sup>1</sup> My obvious response, then, is that those eternal sentences themselves can serve as the truth vehicles. Just think of 'I', 'you', 'he', 'she', 'here', and 'there' as supplanted by names and addresses or other identifying particulars as needed. Think of tenses as dropped; we can use dates, the predicate 'earlier than', and the like as needed. Think of ambiguities and vaguenesses as resolved by paraphrase—not absolutely, but enough to immobilize the truth value of the particular sentence. The truth values need not be known, but they must be stable.

The attitude is the one that is familiar in the teaching of logic. When we take illustrative sentences from everyday language and paraphrase them into the notation of truth functions and quantifiers, we think of the reference of demonstratives and personal pronouns as fixed—albeit tacitly—and we never dream of reading ' $\exists x$ ' as 'there was' or 'there will be something  $x$ '.

Declarative sentences thus refined—eternal sentences—are what I shall regard as truth vehicles in ensuing pages, for the most part. On the whole it is the convenient line for theoretical purposes. We must recognize, though, that it bypasses most of what counts in daily discourse as true or false, since our utterances are not for the most part thus refined. The truth vehicles directly related to behavior are not sentences as repeatable linguistic forms, but rather the individual acts of uttering them. These are for the most part univocal in truth value without benefit of paraphrase. There are just occasional failures, perhaps because some name turns out to be empty or because some vague term turns out to be indeterminate just where it matters for the utterance in question. Such utterances may be dismissed as neither true nor false.

So much by way of coming to terms with the realities of verbal behavior. Let us now return to the more conveniently manageable domain of eternal sentences, whose truth or falsity, known or unknown, is unchanging.

## 2 Truth as Disquotation

Such being what admit of truth, then, wherein does their truth consist? They qualify as true, one is told, by corresponding to reality. But correspondence word by word will not do; it invites the idle cluttering of reality with a bizarre host of fancied objects, just for the sake of correspondence. A neater plan is to posit *facts*, as correspondents of true sentences as wholes; but this still is a put-up job. Objects in abundance, concrete and abstract, are indeed needed for an account of the world; but facts contribute nothing beyond their specious support of a correspondence theory.

Yet there is some underlying validity to the correspondence theory of truth, as Tarski has taught us. Instead of saying that

‘Snow is white’ is true if and only if it is a fact that snow is white

we can simply delete ‘it is a fact that’ as vacuous, and therewith facts themselves:

‘Snow is white’ is true if and only if snow is white.

To ascribe truth to the sentence is to ascribe whiteness to snow; such is the correspondence, in this example. Ascription of truth just cancels the quotation marks. Truth is disquotation.

So the truth predicate is superfluous when ascribed to a given sentence; you could just utter the sentence. But it is needed for sentences that are not given. Thus we may want to say that everything someone said on some occasion was true, or that all consequences of true theories are true. Such contexts, when analyzed logically, exhibit the truth predicate in application not to a quotation but to a pronoun, or bound variable.

The truth predicate proves invaluable when we want to generalize along a dimension that cannot be swept out by a general term. The easy sort of generalization is illustrated by generalization on the term ‘Socrates’ in ‘Socrates is mortal’; the sentence generalizes to ‘All men are mortal’. The general term ‘man’ has served to sweep out the desired dimension of generality. The harder sort of generalization is illustrated by generalization on the clause ‘time flies’ in ‘If time flies then time flies’. We want to say that this compound continues true when the clause is supplanted by any other; and we can do no better than to say just that in so many words, including the word ‘true’. We say “All sentences of the

form 'If  $p$  then  $p$ ' are true." We could not generalize as in 'All men are mortal', because 'time flies' is not, like 'Socrates', a name of one of a range of objects (men) over which to generalize. We cleared this obstacle by *semantic ascent*: by ascending to a level where there were indeed objects over which to generalize, namely linguistic objects, sentences.

Semantic ascent serves also outside of logic. When Einstein propounded relativity, disrupting our basic conceptions of distance and time, it was hard to assess it without leaning on our basic conceptions and thus begging the question. But by semantic ascent one could compare the new and old theories as symbolic structures, and so appreciate that the new theory organized the pertinent data more simply than the old. Simplicity of symbolic structures can be appreciated independently of those basic conceptions.

As already hinted by the correspondence theory, the truth predicate is an intermediary between words and the world. What is true is the sentence, but its truth consists in the world's being as the sentence says. Hence the use of the truth predicate in accommodating semantic ascent.

The disquotational account of truth does not define the truth predicate—not in the strict sense of 'definition'; for definition in the strict sense tells how to eliminate the defined expression from every desired context in favor of previously established notation. But in a looser sense the disquotational account does define truth. It tells us what it is for any sentence to be true, and it tells us this in terms just as clear to us as the sentence in question itself. We understand what it is for the sentence 'Snow is white' to be true as clearly as we understand what it is for snow to be white. Evidently one who puzzles over the adjective 'true' should puzzle rather over the sentences to which he ascribes it. 'True' is transparent.

For eternal sentences the disquotational account of truth is neat, we see, and simple. It is readily extended, moreover, to the workaday world of individual utterances; thus an utterance of 'I have a headache' is true if and only if the utterer has a headache while uttering it.

### 3 Paradox

It seems paradoxical that the truth predicate, for all its transparency, should prove useful to the point of indispensability. In the matter of

paradox, moreover, this is scarcely the beginning. Truth is notoriously enmeshed in paradox, to the point of out-and-out antinomy.

An ancient form of the antinomy of truth is the Paradox of the Liar: 'I am lying', or 'This sentence is not true'. A looser and fancier version was the paradox of Epimenides the Cretan, who said that all Cretans were liars. The underlying antinomy can be purified for logical purposes to read thus:

- (1) 'yields a falsehood when appended to its own quotation' yields a falsehood when appended to its own quotation.

Executing the instructions in (1), we append the nine-word expression to its quotation. The result is (1) itself. Thus (1) says that (1) itself is a falsehood. It is thus tantamount to 'I am lying', but more clean-cut. It hinges only on the innocuous operations of quoting and appending and the notion of falsehood, which reduces to an innocent 'not' and *true*. The truth predicate is clearly the trouble spot. The inevitable conclusion is that the truth predicate, for all its transparency and seeming triviality, is incoherent unless somehow restricted.

For further explicitness a technical turn of phrase will be convenient. The truth predicate will be said to *disquote* a sentence *S* if the form

\_\_\_\_\_ is true if and only if \_\_\_\_\_

comes out true when *S* is named in the first blank and written in the second. Thus what the disquotational account of truth says is that the truth predicate disquotes every eternal sentence. But the lesson of the antinomy is that if a language has at its disposal the innocent notations for treating of quoting and appending, and also the notations of elementary logic, then it cannot contain also a truth predicate that disquotes all its own eternal sentences—on pain of inconsistency. Its truth predicate, or its best approximation to one, must be incompletely disquotational. Specifically, it must not disquote all the sentences that contain it. That was the trouble with (1). And of course it must not disquote all the sentences containing terms by which that predicate could be paraphrased. This, apart from its special orientation to quoting and appending, is substantially what has come to be known as Tarski's Theorem. He has proved harder things.



The truth predicate loses little in general utility thereby, for it can still disquote all the eternal sentences that do not themselves contain it or other expressions to the same effect. And even these excluded applications can be accommodated by a hierarchy of truth predicates. The hierarchy begins with a predicate 'true<sub>0</sub>', which disquotes all sentences that contain no truth predicate or equivalent devices. A predicate 'true<sub>1</sub>', next, disquotes all sentences that contain no truth predicate or equivalent devices beyond 'true<sub>0</sub>'. And so on up. It is a hierarchy of progressively more nearly perfect truth predicates. The plan dates back in a way to the early phase of Russell's theory of types (1908), by which he meant to obstruct the Paradox of the Liar among others.

#### 4 Tarski's Construction

We saw that disquotation is loosely definitive of truth. We may now be thankful for the looseness, seeing as we do that definability of truth for a language within the language would be an embarrassment. And thus it was that Tarski undertook the perilous adventure of defining it for the language within the language, as nearly as possible, if only to see what minimum obstacle saved the situation. This was not his order of presentation, but it comes out the same.

The language chosen for the construction contains the logical notations for quantification and the truth functions and the set-theoretic notation ' $x \in y$ ' for membership.<sup>2</sup> It contains also a finite lexicon, as large as you please, of predicates for natural science and daily life. Finally it contains the means, in effect, of quoting and appending, as in (1); that is, it can specify each of its single signs and it can express the concatenation of expressions.

Truth pertains to closed sentences, that is, sentences without free variables. Its analogue for open sentences is the two-place predicate of *satisfaction*. An assignment of objects to variables *satisfies* a sentence if the sentence is true for those values of its free variables.

What sort of object is an *assignment* of objects to variables? It is simply a function, or one-many *relation*, relating one and only one object to each variable—that is, to each letter, '*w*', '*x*', '*y*', '*z*', '*w*', etc. A relation,

in turn, is a set, or class, or *ordered pairs*. Ways are well known of defining the notation ' $\langle x, y \rangle$ ' of ordered pairs contextually by means of epsilon and the logical particles.

Once satisfaction is defined, truth comes easily; for a closed sentence, having no free variables, is vacuously satisfied by all assignments or none according as it is true or false. We can simply define

(2) ' $y$  is true' as ' $\forall x (x \text{ is assignment} \rightarrow x \text{ satisfies } y)$ '.

So Tarski's big job is to define satisfaction. First he defines it for *atomic* sentences, each of which consists of just a predicate adjoined to one or more variables. For instance an assignment satisfies the atomic sentence ' $x \in y$ ' if and only if what is assigned to the letter ' $x$ ' is a member of what is assigned to the letter ' $y$ '. Correspondingly for each of the other predicates in the lexicon. An assignment satisfies an alternation of sentences, next, if and only if it satisfies one of both of them; it satisfies their conjunction if and only if it satisfies both; and it satisfies a negation if and only if it does not satisfy the sentence that is negated. Finally, an assignment satisfies an existential quantification ' $\exists x(\dots x \dots)$ ' if and only if some assignment, matching that one except perhaps for what it assigns to ' $x$ ', satisfies ' $\dots x \dots$ '.

Such is Tarski's recursive or inductive definition of satisfaction. It explains satisfaction of atomic sentences outright, and it explains satisfaction of sentences of each higher grade or complexity in terms of satisfaction of their components. Universal quantification is passed over because it is expressible in terms of existential quantification and negation in familiar fashion.

## 5 Paradox Skirted

Clearly all the clauses of this inductive definition can be formulated within the formal language itself, except for the word 'satisfies' that is being defined. Thus we have apparently defined satisfaction for the language within the language. Invoking (2), then, we have done the same for truth. This was supposed to spell contradiction.

We could even get contradiction directly from satisfaction, without the detour through (2), 'truth', and (1). We have merely to ask whether

assignment of the sentence 'not ( $x$  satisfies  $x$ )' to the variable ' $x$ ' satisfies the sentence 'not ( $x$  satisfies  $x$ )' itself. Such is Grelling's so-called Heterological Paradox.<sup>3</sup>

What saves the situation is that the definition of satisfaction is inductive rather than direct. The inductive definition explains satisfaction of each specific sentence, but it does not provide a translation of ' $x$  satisfies  $y$ ' with variable ' $y$ '. Consequently it does not translate the 'not ( $x$  satisfies  $x$ )' of Grelling's paradox, and does not support the truth definition (2) for variable ' $y$ '; it just explains truth of each specific closed sentence. It leaves the truth predicate in the same state in which the disquotational account left it; namely, fully explained in application to each specific sentence of the given language but not in application to a variable.

It was a near miss, and I turn now to a nearer one. Treating relations again as classes of ordered pairs, we can write ' $\langle x, y \rangle \in z$ ' to mean that  $x$  bears the relation  $z$  to  $y$ . Now imagine the above inductive definition of satisfaction written out in our formal language, with the variable ' $z$ ' always in place of 'satisfies' and so ' $\langle x, y \rangle \in z$ ' in place of ' $x$  satisfies  $y$ '. Let the whole inductive definition, thus edited, be abbreviated as ' $\Phi z$ '. It fixes  $z$  as the satisfaction relation. Evidently we arrive thus at a *direct* definition:

$$(3) \quad \exists z(\Phi z \cdot \langle x, y \rangle \in z)$$

of ' $x$  satisfies  $y$ ' strictly within the formal language itself. Doesn't this spell contradiction?

No. The catch this time is that there might not be any relation  $z$  such that  $\Phi z$ . Indeed there better not be, on pain, we see, of contradiction. The two-place predicate 'satisfies' remains well defined in its inductive way, but a grasp of the predicate and how to use it carries no assurance of the existence of a corresponding abstract object, a corresponding set of ordered pairs. And, failing such a pair set, (3) fails to translate ' $x$  satisfies  $y$ '. Though the satisfaction predicate is well explained even within the formal language by the recursion, it does not get reduced to the prior notation of that language. Satisfaction, and truth along with it, retain the status that truth already enjoyed under the disquotation account: clear intelligibility without full eliminability.<sup>4</sup>

## Notes

1. In my logic books of 1940, 1941, and 1950, and revised editions down the years, my word for them was 'statement'; but I became chary of it because of its customary use rather for an act. 'Eternal sentence', along with 'standing sentence' (§4), dates from *Word and Object*. 'Standing sentence' is more inclusive. 'The *Times* has come' is a standing sentence, for it can command assent all day independently of interim stimulation; but it is not eternal.
2. Readers expecting a contrast between object language and metalanguage should bear in mind that I am already addressing the aforesaid perilous adventure.
3. See my *Ways of Paradox* (Cambridge: Harvard University Press, 1976), pp. 4–6.
4. The foregoing analysis is adapted from my *Philosophy of Logic* (Englewood Cliffs: Prentice Hall, 1970), pp. 35–46. A somewhat different analysis, in my 1952 paper "On an Application of Tarski's Theory of Truth," is called for when the set theory is of the kind that admits both sets and ultimate classes.



---

## Correspondence Truth, Disquotational Truth, and Deflationism

Hartry Field

### 1 A Theory of Truth Is a Theory of Truth-Conditions

There are two main types of entity that we primarily want to evaluate in terms of truth and falsity: utterances (or sentences, when any ambiguity or indexicality in the sentences can be ignored) and states of thinking.

It is sometimes said that we do not ordinarily apply the word ‘true’ directly to utterances or states of thinking, but rather to some third sort of entity—propositions. But whether this point about English usage is correct seems to me of little interest. Even if it is correct, the point of having a notion of truth applicable to propositions is to facilitate the evaluation of utterances and states of thinking; the point is to enable us to evaluate these things in terms of whether they *express truths* or *have true contents*.

I take the goal of a theory of truth to be to explain what it is for an utterance and/or a state of thinking to be true or false (or if you prefer, to express a truth or to have a true content). This states the goal of a theory of truth independently of any notion of proposition, and that seems all to the good since different people understand that notion in importantly different ways. It may ultimately be helpful or even necessary to introduce propositions into a theory of truth, as theoretical entities; but it would be unfortunate to state the *goals* of the theory in terms of entities whose nature (and even whose existence) is exceedingly controversial.

Note that as I understand a theory of truth, it could equally be called a theory of truth-*conditions*: the question that forms the subject of a theory of truth is substantially equivalent to the question of what it is for an utterance or state of thinking to have a given set of truth-conditions. For

suppose we had an account which told us that for an utterance or a state of thinking to have as its truth-conditions *that Caesar crossed the Rubicon* is for it to have feature Q; then we would know that one way for an utterance or a state of thinking to be true is for it to have Q and for Caesar to have crossed the Rubicon. To say that a given utterance *u* or thought-state *s* is true could be viewed as something like an infinite disjunction of such claims as:

*u* (or *s*) has the truth-conditions that Caesar crossed the Rubicon, and Caesar crossed the Rubicon;

*u* (or *s*) has the truth-conditions that snow is white, and snow is white;

etc. The real philosophical problem (as pointed out in Ramsey 1927) could be put by asking what it is for an utterance or a thought-state to have a given set of truth conditions.<sup>1</sup>

In saying this, I don't mean to be taking a stand on whether we should follow Ramsey in literally defining what it is for an utterance or thought-state to express a truth as an infinite disjunction of claims of the form '*u* (or *s*) has the truth-conditions that *p*, and *p*'. That feature of Ramsey's account requires acceptance of 'substitutional quantification' or some other device of infinite disjunction. If one doesn't accept such a device, one may prefer either to reverse the order of definition by defining 'truth-conditions' in terms of 'true' and a modality (roughly, *S* has the truth-conditions that snow is green iff necessarily it is true iff snow is green),<sup>2</sup> or to define both 'true' and 'truth-conditions' recursively in terms of some related notions such as reference (as in Field 1972, for instance). The order of explication seems to me a matter of technical detail that is of no great metaphysical substance. Whatever the order of explication, Ramsey's remark about where the real philosophical problem lies seems apt.

## 2 Pure Disquotational Truth

A prevalent view in the early days of the Vienna Circle was that the notions of truth and of truth conditions are a piece of useless metaphysics that we ought to abandon. For instance, legend has it that Neurath put 'true' on an 'Index of Prohibited Words'. Ayer's view in chapter 5 of

*Language, Truth and Logic* was slightly less radical, but quite similar in spirit: he proposed that we allow talk of our utterances being true and having truth conditions, but that we give a deflationary account of such talk. Indeed, this feature of Ayer's view is closely tied to his verifiability theory of meaning: it is essential to a verifiability theory of meaning and thought to dethrone truth-conditions from the central place in the theory of meaning and the theory of thought that they had had in the work of Frege, Russell, early Wittgenstein and Ramsey.

The last name on this list may seem surprising, since when Ayer (1936) describes his 'redundancy theory of truth', he claims to be *following* Ramsey. In fact however Ayer's and Ramsey's views are in total opposition. One way to express the difference between Ayer and Ramsey is to say that Ramsey assumed that utterances and thought-states have *correspondence* truth conditions; whereas Ayer assumed that they do not, but have only *disquotational* truth-conditions. I will need to explain what this means.

As a preliminary to doing so, let's ask why we need a notion of truth. According to W. V. Quine (1970, chapter 1) and Stephen Leeds (1978), we need it for purely logical reasons: 'true' is simply a device for infinite conjunction (or disjunction), a device which would be of little use were we to have in our language other devices of infinite conjunction.<sup>3</sup>

To illustrate why a device of infinite conjunction is useful, let's look at axiomatic theories. Often when we formulate theories rigorously in first order logic, it is essential to use infinitely many axioms. This is generally regarded as acceptable as long as all of these axioms are obtained from a small number of axiom schemas (together with a small number of separate axioms)—schemas like, for instance, the schema of mathematical induction or the completeness schema in Euclidean geometry. In that case one can in a certain sense explain what it is to accept the theory without introducing a notion of truth—roughly, to accept the theory is to accept each of the separate axioms and be disposed to accept each instance of the axiom schemas. But the idea of *rejecting* such a theory is more problematic, given that one can reject a theory on the grounds that it has unacceptable consequences without knowing which part of the theory ought to go. If a theory is finitely axiomatized, then we must say not that we reject some axiom, but that we reject the conjunction of the axioms;



but if the theory is not finitely axiomatized, then there *is* no conjunction of the axioms in any straightforward sense. But as a surrogate for an infinite conjunction, we introduce the term ‘true’ and semantically ascend, saying ‘Not every axiom of this theory is true.’

The example of denying an infinitely axiomatized theory is simply one example of why a surrogate for infinite conjunction is useful; there are lots of other examples as well, and they do not arise only in mathematics. Indeed, I think that in ordinary conversation most of the uses of ‘true’ that prove difficult to paraphrase away are uses where it is serving as a surrogate for infinite conjunction of infinite disjunction. The same holds for a great many *philosophical* uses of ‘true’ as has been pointed out in Grover, Camp and Belnap (1975, pp. 120–1) and Soames (1984, section I). For instance, consider the philosophical claim that there are true sentences that no one will ever have any grounds for accepting. ‘True’ here is naturally understood as a device of infinite disjunction: the claim says that *either* the maximum number of brontosauruses that ever existed at one time was 10,732, but no one will ever have grounds for accepting that; *or* that the total amount that Frank Sinatra has spent on shirts in his life is exactly \$86,526.33, but no one will ever have grounds for accepting that; *or* ...

Now, it is not difficult to imagine other devices than ‘true’ that could be used to express the relevant infinite conjunctions and disjunctions. One device which expresses the relevant infinite conjunctions or disjunctions directly (without the need to ‘semantically ascend’ and talk of sentences) is a substitutional quantifier (with formulas and sentences among the substituends). So if ‘true’ is simply a device of infinite conjunction, then we have a serious need for a predicate of truth only because (or, only *if*) we don’t have a substitutional quantifier in English.<sup>4</sup>

In fact, a substitutional quantifier can be used to define a certain notion of truth: ‘*x* is true’ is defined as the infinite conjunction of the sentences ‘if *x* is “snow is white” then snow is white’, ‘if *x* is “grass is green” then grass is green’, and all similar sentences.<sup>5</sup> This notion of truth (which is of course just the notion motivated by the point about denying theories that are not finitely axiomatized) might be called, following Quine, a notion of *disquotational truth*. Among its characteristic features are that:

- a. One can meaningfully apply it only to sentences that one understands (or can mean something by),
- b. the property of those sentences which it attributes is one that a sentence has or fails to have *independently of the way that the sentence is used by speakers*.

Concerning (a): the point isn't merely the rather trivial one that we don't normally know whether to apply 'true' to sentences what we don't understand (or can't mean anything by); it is that it makes no sense to wonder whether such sentences are true in the disquotational sense. To attribute disquotational truth to a sentence is cognitively equivalent to uttering the sentence, so if one doesn't understand (or can't mean anything by) the sentence, then one doesn't understand (or can't mean anything by) the attribution of disquotational truth. To illustrate (b): the sentence (C<sub>1</sub>) if we had used the word 'white' differently, 'grass is white' might have been true

is equivalent (if 'true' is used disquotationally) to:

- (C<sub>2</sub>) if we had used the word 'white' differently, grass might have been white

by the complete cognitive equivalence of "'Grass is white" is true' and 'Grass is white'. In other words, C<sub>1</sub> comes out false if 'true' is understood disquotationally.

This is not meant as a criticism of the notion of disquotational truth. It would be a criticism if there were no way to express in terms of disquotational truth the gist of what someone is trying to convey in asserting C<sub>1</sub>; but there are ways to do this, as I will discuss shortly. Indeed, it seems to me a virtue of the notion of disquotational truth that it has feature (b), i.e. that it makes C<sub>1</sub> false: a notion of truth for sentences that *didn't* have this feature wouldn't well serve the purpose that Quine and Leeds have identified. That is, I take it that a notion of *correspondence truth* (applicable to sentences or utterances) is, among other things, a notion of truth which differs from disquotational truth in making C<sub>1</sub> and C<sub>2</sub> inequivalent; if so, then if I want to deny Euclidean geometry, I don't want to express my denial by saying that not all axioms of Euclidean geometry are true in a correspondence sense. For what I want to say is something about the structure of space only, not involving the linguistic

practices of English speakers. Admittedly, a denial that the axioms of Euclidean geometry are all true in a correspondence sense could be used to *convey the belief that* they are not all disquotationally true (i.e. the belief in the denial of their infinite conjunction); for the relevant facts about English might be common knowledge, and modulo these facts of common knowledge, correspondence truth and disquotational truth are the same. The fact remains however that the belief that we are trying to convey does not involve correspondence truth. We see then that even someone who accepts a notion of correspondence truth needs a notion of disquotational truth (or some other means of expressing the relevant infinite conjunctions, e.g. a substitutional quantifier) in addition. (Propositions could be used to serve the role, but disquotational truth serves it more cheaply.)

### 3 Deflationism, Ramsey and Ayer, Extended and Modified Disquotational Truth

I have argued that ‘correspondence truth’ (whatever exactly that is) is ill-suited to serve the purposes that disquotational truth serves. In that case, what purposes *does* it serve? I take it to be the core of Neurath’s and Ayer’s view—and more recently, Quine’s and Leeds’s—that the answer is that it serves no useful purpose at all, and hence that theorizing about correspondence truth is pointless at best. Any view that adheres to this position while at the same time (contrary to Neurath) preserving a use for the word ‘true’ will be called a *deflationary* conception of truth.<sup>6</sup>

Deflationism as so characterized is only as clear as the notion of correspondence truth; and at this point any way, that is not very clear at all. I trust that the reader has *some* sense as to what people have meant by correspondence theories of truth; and a correspondence notion of truth is, of course, a notion of truth for which a correspondence theory provides a correct account. We can partially elaborate the familiar conception, I think, by saying that correspondence truth differs from disquotationalist truth in not having properties (a) and (b) from the previous section, and in making counterfactuals  $C_1$  and  $C_2$  come out inequivalent. More generally, correspondence truth-conditions are supposed to be objective features of an utterance or thought-state, features which the utterance or

thought-state could have whether or not we know it, and which it has in virtue of facts about the relations between the utterance or thought-state and the world around us; also, of course, correspondence truth-conditions are such that from truth-conditions together with 'the facts' one can immediately determine truth-values. This is all pretty vague, but I think that in order to provide a clear positive characterization of what a correspondence theory or a correspondence notion of truth is, we would need to first get clear on what if any point there might be to having a notion of truth that goes beyond the disquotational. Since that is the main question to be discussed in this paper, it is impossible to give a clear positive characterization of correspondence truth at the start.

For now, then, I prefer to try to characterize a correspondence notion of truth negatively. I have already explained what a purely disquotational notion of truth is. This enables me to explain what *one form* of deflationist position is: this form of deflationism holds that *the only useful notions of truth are disquotational truth and various other notions definable from it using rather limited additional resources*. There are various other possible deflationary positions, theories that try to interpret 'true' in epistemological terms like coherence with observations or warranted assertibility, but I will not be taking those positions seriously. Putting such epistemological views of truth aside, what I mean by a correspondence notion of truth is simply a non-epistemological notion of truth that is not explainable in terms of disquotational truth (or any other purely logical notion of truth) using only limited additional resources.

This is still pretty vague, of course, for I have not said anything about what additional resources count as 'limited'. (Note also that I have not claimed that *any* notion of truth that, when applied to our own utterances, violates (a) and (b) above is a correspondence notion; for I have not ruled out that a deflationist might construct a notion with these features from disquotational truth plus 'limited resources'. More on this possibility at the end of this section.) Despite the vagueness of my characterization, I think that there is little difficulty in classifying many views of truth in terms of whether or not they are deflationary.

One view which is clearly *not* deflationary is Ramsey's: Ramsey is a clear case of a correspondence theorist. To be sure, Ramsey did not think that *the word 'true'* expresses a correspondence notion: he held that this

word applies primarily to what he called ‘propositions’, which are simply encapsulations of truth-conditions; and that so applied, it is redundant, in the sense that to say of the proposition that Caesar crossed the Rubicon that it is true is simply to say that Caesar crossed the Rubicon. But Ramsey recognized that the problem of truth in my sense (the sense indicated in my third paragraph) was not thereby trivialized: he concluded that ‘the real problem is not as to the nature of truth and falsehood, but as to the nature of judgement or assertion ...’ (1927, 143). That is, the real problem concerns what it is for an utterance or thought-state to *express the proposition that* Caesar crossed the Rubicon, i.e. to *have the truth conditions that* Caesar crossed the Rubicon.

On Ramsey’s view, ‘expresses a true proposition’ is clearly a correspondence notion (i.e. not a deflationist notion). This is clear from the account that Ramsey gives of what it is for a thought-state to express the proposition that Caesar crossed the Rubicon. Such a thought-state must have ‘mental factors’ which Ramsey takes to be ‘words, spoken aloud or to oneself or merely imagined’ (p. 144); to think that Caesar crossed the Rubicon is to be in a state of thinking that connects up (in the right way) a mental word for Caesar with a mental word for the relation of crossing and a mental word for the Rubicon.<sup>7</sup> What exactly it is for a word to stand for Caesar, or the relation of crossing, or whatever, is left unanalysed, but it is clear that Ramsey thought that some sort of naturalistic explication of what it is for a mental symbol to stand for an object or relation was appropriate. Clearly on this view ‘expresses a true proposition’ does not behave at all like disquotational truth or like any notion definable from that with minimal resources.<sup>8</sup>

Despite the citation of Ramsey in Ayer (1936), I think it is clear that what Ayer had in mind was quite different from this. One indication that this is so is provided by the whole tenor of Ayer’s discussion: Ayer suggests not that the problems which some have found with the notion of truth be relocated as problems about what it is to utter a given proposition or to think a given proposition, but rather that the only real problem in this area is one of *verification*. Another indication is the fact that Ayer defines ‘proposition’ not in Ramsey’s way (where propositions have entities like Caesar and the Rubicon and the relation of crossing as constituents), but as ‘a class of sentences which have the same intentional

significance for anyone who understands them' (1936, 88).<sup>9</sup> Given this account of propositions as equivalence classes of sentences, the problem of explaining the relation between on the one hand an utterance or thought-state and on the other hand a proposition is bound to be quite different and rather less substantive than it was for Ramsey. Indeed, Ayer's view seems to have been that for each English sentence *p*, the phrase 'the proposition that *p*' is equivalent in meaning to 'the equivalence class that contains the sentence "*p*"'. Ayer doesn't use the phrase 'truth-conditions', but it would be natural on his view to use it too for an equivalence class of sentences that includes sentences of our own language. So to assert that 'Queen Anne is dead' expresses the proposition that Queen Anne is dead, or that it has the truth-conditions that Queen Anne is dead, is not to say anything substantive about the use of the sentence 'Queen Anne is dead': the sentence would have had the truth-conditions that Queen Anne is dead, however the sentence was used. We have, then, a *disquotational* notion of truth-conditions, in that the analogue of clause (b) in my explanation of the notion of disquotational truth applies (at least as regards the truth conditions of one's own utterances).

Clause (a), of course, does not apply: on Ayer's view it makes sense for me to speak of an utterance in a language I don't understand as expressing the proposition that (or having the truth conditions that) Queen Anne is dead:<sup>10</sup> what this means is that that utterance has 'the same intensional significance' as my utterances of 'Queen Anne is dead'; and that in turn can be explicated *without relying on anything like a notion of truth*, according to Ayer, for it just means that that utterance and mine have the same verifiability conditions.<sup>11</sup> So we have really an *extended disquotational* notion of truth and truth-conditions, one obtained from a purely disquotational notion (applicable only to sentences I understand) by what Ayer regarded as a rather limited additional device of interlinguistic synonymy. Of course, there are other advocates of deflationary views of truth, notably Quine, that don't regard this additional device as so harmless, and who consequently would view with suspicion any such extension of the notion of disquotational truth to languages one doesn't understand. On Quine's view it makes perfectly good sense for me to ask whether a foreigner's remark (or a remark by

another English speaker) is true *as I interpret what the speaker meant*, i.e. *as I would translate his remark into my language or preferred idiom*.<sup>12</sup> But if I can't interpret his remark, it makes no sense to inquire whether it is true (except perhaps in counterfactual terms about how I would interpret it under definite conditions); and even if I can interpret it, it makes sense only to ask whether it is true *as I interpret it*, not *as it was really meant*. Quine's arguments for viewing talk of translation with suspicion seem to me strong once one grants the deflationist premise, but this is not the place to adjudicate them.

If one holds that *sameness of meaning* can be defined in terms of *meaning of* in the obvious way (two sentences have the same meaning if the meaning of one and the meaning of the other are identical), and that *meaning of* can be explicated without reliance on any truth-theoretic notions, one might want to introduce a *modified disquotational* notion of truth-conditions that differed from pure disquotational truth conditions in making counterfactuals  $C_1$  and  $C_2$  inequivalent, as well as in regard to feature (a). For instance, if the meaning of a sentence is just its verification conditions, then we could define ' $x$  has as its truth-conditions (in the modified disquotational sense) that snow is white' as:

the verification conditions of  $x$  are the same as the verification conditions that 'snow is white' *actually* has;

the idea of the 'actually' operator being to force a comparison with 'the actual world' even inside modal and counterfactual contexts. Something analogous could be done for other, more plausible, explications of what interpersonal sameness of meaning might consist in: e.g., '(interpersonal) sameness of conceptual role'.<sup>13</sup> (The definition of modified disquotational truth depends on treating verification conditions or conceptual roles or 'meanings' as explicable independently of *sameness of meaning*. So the modified disquotational position not only requires one to put aside Quinean doubts about interpersonal sameness of meaning, it also requires us to put aside *further* Quinean doubts about explaining meaning as something other than equivalence classes under the sameness of meaning relation. Perhaps these further Quinean doubts are not so serious, though.)

Such a modified disquotational notion would not well serve the purposes that a purely disquotational notion serves: e.g. we'd still need the purely disquotational notion if we wanted to deny an infinitely axiomatized theory like Euclidean geometry without mentioning the verification-conditions of English sentences. What purposes would such a notion serve? Well, if according with ordinary intuitions about counterfactuals counts as a purpose, it serves that: if 'true' is read as modified disquotational truth, our counterfactual  $C_1$  comes out true. But it seems to me that we really don't need the controversial semantic assumptions that go into the notion of modified disquotational truth to alleviate the apparent counterintuitiveness of declaring  $C_1$  false, for the gist of what someone intends to assert in saying  $C_1$  is captured by the claim that in thinking about imaginary circumstances in which our language is used differently, it is natural to semantically evaluate this language relative to a mapping into itself other than the identity mapping, and 'Grass is white' might be true relative to such a map. This can be said by someone skeptical of interpersonal synonymy, or rather interworld synonymy, for no claim need be made that the mappings one considers reflect an objective synonymy relation. But if meanings or interworld synonymy can be made sense of, modified disquotational truth does seem like a natural notion to employ. And since it mimics correspondence truth so well in its modal properties, this raises the question of whether it mightn't be a good surrogate for correspondence truth. That is, perhaps it will turn out that whatever purposes a full-blooded correspondence notion of truth has been thought to serve would be served equally well (or better) by the modified disquotational notion. If this turned out to be the case, and if all the Quinean doubts about the modified disquotational theory could be shown to be misplaced, including especially those that arise for the extended disquotational notion as well, we should probably regard that as a vindication of the deflationist position.

My own view, however, is that this is not the best way for a deflationist to try to vindicate his deflationism. In my opinion, a *radical* deflationism that proposes that no serious use be made of either extended or modified disquotational truth or truth-conditions is to be preferred both to a *weak* deflationism that relies on modified disquotational truth and to a *moder-*



*ate* deflationism that relies on extended but not modified disquotational truth. (Whether radical deflationism is to be preferred to a correspondence theory is a different question, and I think a more difficult one.) I prefer radical deflationism to weak and moderate deflationism for two reasons:

- i. I am inclined to share Quine's doubts about intersubjective sameness of meaning, especially in so far as it is to be explicated independently of correspondence truth conditions;
- ii. I doubt that extended or modified disquotational notions of truth conditions could be put to much use anyway.

I do not discuss (i) in this paper, but there is some discussion of (ii) in the longer version (Field 1986) from which what appears here has been excerpted.

The reader may wonder what the conception of a correspondence notion could be, that makes the modified disquotational notion of truth not count as a correspondence notion. (After all, the notion is like correspondence notions as regards features (a) and (b).) Is it that no theory worthy of the name 'correspondence theory' could allow that verification-conditions (for non-indexical sentences anyway—see note 11) determine truth-conditions (in the sense that any two non-indexical sentences with the same verification-conditions have the same truth-conditions)?<sup>14</sup> No, I see no motivation for restricting the term 'correspondence theory' like that. Rather, the view is that any theory worthy of the name 'correspondence theory' which holds that verification conditions for non-indexical sentences determine truth-conditions (and hence, in conjunction with the appropriate facts, determine truth-*values*) must ultimately offer an account of *how* they determine truth-conditions (and truth-values): it can't just appeal to the disquotation schema to set up the correlation. An analogous condition is of course to be imposed on a theory that holds that truth-conditions are determined by something other than verification-conditions, e.g. by 'conceptual roles'. I choose to accept this additional condition on correspondence theories (i.e. this condition that goes beyond the rejection of (a) and (b)) partly because it seems intuitively natural to do so, but partly for a more theoretical reason. For it seems to me that *if* there is an argument to be made for going beyond disquotational or extended disquotational truth-conditions to something like correspondence truth-conditions, this argument motivates only those

theories that meet the extra condition: modified disquotational theories would not be enough. And if there is no such argument to be had, if a full-blooded correspondence notion of truth can be rejected, then I see no value in a modified disquotational notion of truth (unless one attaches value to disguising the radicalness of one's position). It would be more up front to declare that one recognized no notion of truth beyond a purely disquotational one, and no semantic notions beyond verification conditions or conceptual role or whatever.

#### 4 Tarski's Theory of Truth

How does Tarski's work on truth relate to correspondence and disquotationalist notions?

It would be natural to say that Tarski made two main contributions to the theory of truth.

Tarski's first contribution was to show how purely disquotational truth (or what is almost the same thing, substitutional quantification with formulas as the substitution class) is definable in purely first-order logic, provided that sufficient ontology is available. Suppose that *L* is a fragment of our own language—where by 'our own language' I mean the totality of sentences that we understand. Then the obvious ('homophonic') Tarskian definition of truth-in-*L*, in our own language, is essentially the same as a purely disquotational notion of truth: it is a device of infinite conjunction. (I will make some qualifications on this claim in note 16 below.)<sup>15</sup> In Field (1972) I made a mistake in underestimating the value of a disquotational truth-predicate, and consequently failed to see that the task of defining it without a substitutional quantifier might be important. The value of the accomplishment can still be debated—there is a question of whether the use of ontological resources to buy a reduction in logical resources is appropriate—but that is a matter I do not intend to pursue here.<sup>16</sup>

Tarski's second contribution was to define a relativized notion, namely 'true in a model'. Like disquotational truth, this is a mathematical notion, and is of course available to the deflationist. But 'true in a model', unlike the disquotational notion of truth, is applicable to languages that we don't understand as well as to languages that we do: it is not a dis-

quotational notion.<sup>17</sup> Another reason that it is not disquotational is that in the purely disquotational sense, truth is unrelativized. That is, my own sentence ‘Caesar crossed the Rubicon’ is true—not in a model, but just true—if and only if Caesar crossed the Rubicon. The fact that in some models the sentence is true if and only if Columbus swam the Nile does not affect this, for purely disquotational truth and truth in a model are totally different things. ‘True in a model’ is likewise not an extended disquotational notion or a modified disquotational notion, for it is defined without reference to any notion of sameness of content. ‘True in a model’ is not a disquotational notion at all. But as I’ve said, it is available to the deflationist (radical, moderate or weak) nonetheless. (And it is a notion with fairly uncontroversial uses, e.g. in defining logical consequence, logical truth, and the like.)

We might modify Tarski’s definition of ‘true in a model’ to a recursive definition of ‘true in an interpretation’, where an interpretation is just like a model except that quantifiers are not confined to ranging over some definite set but are allowed to range over absolutely everything. (It will assign objects to individual constants, sets to 1-place predicates, etc.) (If one wants, one can alter this for the case of mathematical predicates like ‘ $\epsilon$ ’ that don’t have sets as their *disquotational* extension, so that there could be some interpretation (‘the homophonic interpretation’) such that truth in that one interpretation corresponds to disquotational truth.<sup>18</sup> I will not bother to work out what the best modification would be.) Like ‘true in a model’, ‘true in an interpretation’ is not a disquotational notion (pure, extended or modified); and there is no way to *explicitly* define it in a first-order language. But we should probably regard it as available to a deflationist nonetheless.

If the notion is available to the deflationist, it is also available to the correspondence theorist; and one way of trying to develop a correspondence theory is to use the notion of ‘true in an interpretation’ together with the supplementary idea of an interpretation being ‘*relevant*’ or ‘*intended*’, where it is assumed that the notion of ‘relevance’ or ‘intendedness’ plays an important role in understanding the workings of the language. Suppose one could give a substantive account of what it is for an interpretation to be ‘relevant’ to the semantics of a speaker’s language; then one could define an unrelativized notion of truth by defining ‘S is true’ as ‘S is

true in the relevant interpretation of the language'.<sup>19</sup> (This assumes there to be a uniquely relevant interpretation. If there can be more than one interpretation that comes out 'relevant', we could alter the definition to 'S is true in all relevant interpretations'; or we could keep to the original definition and hold that there is no fact of the matter as to whether a sentence that is true on one relevant interpretation and false on another comes out true.) Should 'true', so defined, count as a correspondence notion? That depends on what sort of account we give of the notion of relevance. If we define relevance by saying that an interpretation J of our own language is relevant if and only if, for any *p*, '*p*' is true in J if and only if *p* (understanding the quantification as substitutional), then we have defined relevance (via substitutional quantification) in such a way that the resulting notion of truth for our language is purely disquotational. And if we say that an interpretation J of a language we don't understand is relevant iff for any S of that language and any sentence '*p*' of ours that matches S in conceptual role, S is true in J if and only if *p*, then the resulting notion of relevance and of truth-conditions is extended disquotational. But if we explain what it is for an interpretation of any language, *even our own*, to be relevant *in terms of the way that speakers use the language* (e.g. if we require of a relevant interpretation that it assign to one of our names an object that is causally connected to that name, as in a causal theory of reference), and if we go on to show that this notion of relevance plays an important role in the workings of the language (or in the explanation of the behaviour of the users of the language), then we are viewing relevance and hence truth as correspondence notions. So a radical deflationist cannot allow that any notion of 'relevant interpretation', applicable to other languages as well as our own, plays an important role in the workings of language (or in the explanation of the behaviour of language users): for a radical deflationist is debarred from both the correspondence construal and the extended (or modified) disquotationalist construal of the notion.

Philosophers such as David Lewis (1969 and 1975) and Scott Soames (1984) like to say that what Tarski did is define 'true in L' where L is an *abstract interpreted language*, i.e. an ordered pair of a syntax and a model or interpretation. If languages are so conceived, a notion of truth-in-L is not essentially different from the notion of truth in a model or

interpretation: it is neither a disquotational notion nor a correspondence notion, but rather a purely mathematical notion to which either a radical deflationist or a moderate or weak deflationist or a correspondence theorist can perfectly well appeal. In particular, such a notion of truth-in-L may be usable in a correspondence theory of truth: what is required is that we connect the abstract languages up somehow with utterances or thought-states, via a notion like ‘relevance’. Those who like to use Tarskian truth-predicates so conceived usually connect such notions up with utterances: they think that we can make sense of the notion of *the (interpreted) language employed in a given population*. For such philosophers, the notion ‘utterance *u* is true in the interpreted language employed by its utterer’ is a notion that applies to utterances in languages that we don’t understand as well as languages that we do, and it expresses speaker-dependent features of those utterances; indeed, on my understanding of Lewis’s and Soames’s conception of what it is to employ a language (their analogue of ‘relevance’), this notion seems like a full-fledged correspondence notion. (See, for instance, Soames 1984, 428 [chap. 17 of this volume—Ed.]. When Soames claims on p. 429 to be offering a ‘deflationary’ conception of truth, then, he is being misleading.)

There are advantages and disadvantages to introducing interpretations (or abstract interpreted languages) in this way and putting a correspondence theory into this format (as opposed, e.g., to Ramsey’s format involving primitive substitutional quantification, or to the format of Field (1972) where I avoided abstract languages and used only a modified version of Tarskian predicates). That is a matter for debate within the camp of correspondence theorists. My point here is simply this: *the advocacy of ‘a Tarskian notion of truth’ takes on a very different significance if one is thinking of the languages to which it applies as abstract entities and is regarding the notion of ‘employing’ such an abstract entity as a legitimate notion, than it does if one is simply using an unrelativized Tarskian predicate for a language one understands as a device of infinite conjunction*. (And it takes on still different significance if one simply utilizes the relativized notion of truth in a model or interpretation or ‘abstract language’, without utilizing any notion of such an interpretation or abstract language being ‘relevant’ or ‘intended’ or ‘employed’.) Here we have another example<sup>20</sup> of how in the theory of truth a super-

ficial similarity between views often masks fundamental disagreements, and how superficial differences between views often mask fundamental agreement.

## Notes

This chapter is excerpted, with changes, from Field 1986.

1. This way of putting 'the real philosophical problem' may be thought to introduce some difficulties—e.g., the term 'truth-conditions' itself is used in slightly different ways by different philosophers (though the variation is not nearly as great as with 'proposition'). But my point in mentioning the formulation of the problem in terms of truth-conditions is not to find a way to improve the original formulation (the one in terms of truth), but to point out that for those who are comfortable with the second formulation it is not substantively different from the first.
2. This would need fixing up or at least clarifying, to avoid the objection that even for the sentence 'snow is green' the necessity claim won't hold, since that sentence (like all others) has the truth-conditions it has only contingently.
3. I think that Quine and Leeds would acknowledge that it would still have *some* point. For instance, it would enable us to say 'that's true' instead of repeating what someone said; and as Strawson (1950 [chap. 19]) and Grover, Camp and Belnap (1975) have emphasized, this is a simple way to repeat what someone said while simultaneously acknowledging our unoriginality.
4. Grover *et al.* give an added wrinkle to this: they hold that we do have a substitutional quantifier in English, but that it is expressed using the word 'true' (where 'true' does not serve as a predicate but as part of what they call a 'prosentence').
5. 'All similar sentences' ought to include many sentences that contain 'true', but not those sentences which, in conjunction with empirical facts, would lead to semantic paradox if they were included. One can give a version of substitutional quantification which yields this result: the metatheory for the substitutional quantifier will be modelled on the metatheory for a truth-predicate given by Kripke (1975). (Kripke notes the possibility of such a version of substitutional quantification in his note 31.)
6. I take the term from Horwich (1982), though there may be differences between my conception of deflationism and his.
7. This is not strictly accurate: what Ramsey actually says is that this would be true if Caesar, the Rubicon and the relation of crossing were 'simple in relation to [the thinker's] language' (p. 145). Presumably his view was that words like 'Caesar', 'crosses', etc. were defined from more basic mental symbols, but that atomic thought-states involving the more basic symbols worked in the way here described.
8. Another view which may be rather less deflationary than it first appears is the view offered by Grover, Camp and Belnap (1975): notice the free use they make

of notions like *saying that p* and *believing that p*; and notice the remark they make about content and meaning in the paragraph on pages 113–14 of their paper. I think in fact that, because they say nothing about their conception of saying that or believing that, their paper is neutral on the issues involved in the debate about deflationism; though certainly their apparatus of prosentences is one that a deflationist might well want to employ (e.g. in getting the effects of substitutional quantification in a linguistically natural way).

9. Ayer is, of course, ignoring the existence of ambiguities and indexical elements in our sentences; and to avoid introducing boring complexities, I shall follow him in this practice. If one wants to take either ambiguity or indexicality into account, one must view propositions as equivalence classes of *utterances* rather than of sentences. In addition, taking indexicality into account would require that one understand ‘same intentional significance’ in such a way that different people’s utterances of ‘I am sick’ differ in intentional significance; more generally, in such a way that ‘intentional significance’ is dependent on meaning and context together (where ‘context’ includes utterer and place and time of utterance).

10. Maybe I can’t rationally *assert* that the sentence has those truth-conditions without understanding the sentence; but I could *consider the consequences of assuming* that the sentence had those truth-conditions.

11. Recall that I am ignoring complications raised by indexicals here. When such complications are not ignored, we must first decide on how the term ‘truth-conditions’ is to be used for indexical sentences. If we decide to use it in such a way that ‘I am sick’ has different truth-conditions for different utterers and times of utterance, then we must formulate Ayer’s view as the view that sameness of verification conditions determines not sameness of truth-conditions, but rather something like sameness of function from context of utterance to truth-conditions. If, on the other hand, we use ‘truth-conditions’ so that indexical sentences keep the same truth-conditions in different contexts, no such change in the statement of Ayer’s view is needed, but we must instead complicate the relation between truth-conditions and truth. Which course to choose is of course a purely verbal issue. Analogous complications about the treatment of indexicality arise for the conceptual role theories mentioned below.

12. Note also that the *motivation* for a purely disquotational notion of truth works when I apply the word ‘true’ to languages I understand, even if those languages are different from my main language (and from the language in which the word ‘true’ is normally used). If I understand ‘*der Schnee ist weiß*’ then I can be interested in infinite conjunctions involving it. And presumably I can use a disquotational ‘true’ to express such infinite conjunctions. For presumably I *understand* “‘*der Schnee ist weiß*’ is true if and only if *der Schnee ist weiß*”, even if such a mixing of languages is a bit unusual.

13. Interpersonal sameness of conceptual role is supposed to be, like sameness of verification conditions, a purely evidential notion, hence explicable without reliance on truth (or other truth-theoretic notions such as reference or truth-conditions or functions from contexts to truth-conditions). It is supposed to differ

from sameness of verification-conditions in various ways which will make it substantially more plausible as an explication of synonymy: e.g. the conceptual role of a sentence is to take account not only of its evidential relations to observation sentences, but of its evidential relations to other sentences as well. I will not discuss here whether the well-known difficulties in defining a reasonable notion of inter-subjective sameness of conceptual role can be overcome. (For some inconclusive remarks on the pessimistic side, but coupled with claims about the importance of a notion of *intrasubjective* sameness of conceptual role, see Field 1977, 1978. For some remarks on the optimistic side, see Loar 1981 and Block 1985.)

14. Such a determination of truth-conditions by verification-conditions, generalized so as to cover indexical sentences, might be argued to be refuted by 'Twin Earth' phenomena (Putnam 1975) and 'arthritis' phenomena (Burge 1979). But perhaps such phenomena can be assimilated to cases of indexicality.

15. One qualification can be made now: Tarski himself thought that there couldn't be a *single* disquotational truth-predicate; he thought that in order to avoid semantic paradoxes we must make do with a hierarchy of broader and broader truth-predicates 'true<sub>0</sub>', 'true<sub>1</sub>', 'true<sub>2</sub>', etc. Kripke has developed a method for avoiding Tarski's stratification of the truth-predicate (Kripke 1975).

16. Two substantial qualifications are needed on the claim that a homophonic Tarskian truth definition (or its Kripkean variant—see note 15) defines a disquotational truth-predicate. First, what Tarski defined is equivalent to disquotational truth only *modulo standard mathematics*: it cannot be regarded as logically equivalent to disquotational truth (unless standard mathematics is taken as not only true but *logically* true); and it is not usable as a surrogate for disquotational truth in contexts where an attitude of less than certainty is taken toward standard mathematics. Second, Tarski does not really provide a method for explicitly defining the disquotational truth-predicate for our language in our language, he only defines the disquotational predicate for a fragment of the language (obtained by restricting the quantifiers to range over only the members of some specific set). To put it another way, *what is explicitly defined is not really 'disquotationally true', but rather 'comes out disquotationally true under a restriction of the quantifiers to elements of the set D'*. To be sure, the restriction of the quantifiers is needed only at the stage where Tarski turns a recursive definition into an explicit definition: this means that if our goal is only to recursively define 'true', then Tarski has shown how to do it. It might be felt that recursively defining a concept is sufficient to give it clarity, even in absence of an explicit definition; I suspect however that this feeling derives from an acceptance of a notion of disquotational truth or substitutional quantification, and so is not applicable when it is disquotational truth or substitutional quantification that is being recursively defined.

This second difficulty (like the first) arises equally or more for the claim that Kripke's variant of Tarski *defines* a disquotational notion of truth for our language. In some ways, of course, Kripke's definition of truth is better than Tarski's at capturing the disquotational notion: it applies to sentences that themselves contain 'true' (and does so in a way that avoids paradox, by allowing that



sometimes neither a sentence nor its negation be true). But it too requires the restriction of the quantifiers to a set *D*, in order to obtain even the explicit definition of the initial approximation to the final truth definition. And the restriction of the quantifiers to a set is put to further use in Kripke's construction, e.g. in the fixed point theorem (at least in the case where it is satisfaction rather than truth that is being defined—really the case of interest); so here we do not even have a recursive definition of 'true' as opposed to 'comes out true when the quantifiers are restricted to set *D*'. (This point is not intended to downgrade the interest of Kripke's work for understanding disquotational truth. The work is interesting because it casts considerable light on deciding what are reasonable axioms to impose on a disquotational truth-predicate (or on a generalized substitutional quantifier of the sort indicated in note 5).)

Either of the two points made in this footnote seem to me sufficient to show that we would do better to think of disquotational truth as defined from primitive substitutional quantification than to view it as defined *à la* Tarski.

17. I do not say that we have a single notion of truth in a model applicable to *arbitrary* languages. Rather, we have various fairly general notions, each applicable to all languages that have the same 'semantic categories'. A list of the 'semantic categories' (e.g. individual constant, predicate, etc.) employed in the languages in question is needed to define the notion of model that is appropriate to those languages. But though the notion of truth in a model is not completely general, it is general in the sense that it is applicable to more than one language. And, of more importance here, we needn't understand the words in the language to apply it, we need only know what semantic categories the language employs.

18. The idea that disquotational truth is truth in one specific interpretation, the homophonic one, needs a qualification anyway: some words in our language, like 'bald', have no determinate extensions; 'the homophonic interpretation' is not uniquely specified as regards them. An analogous point holds for more interesting cases where we discover to our surprise that one of our words has no determinate extension. Cf. the case of 'mass', discussed briefly in Field 1973. So even when dealing with words that *seem to us* quite determinate, no one 'homophonic interpretation' of them may be uniquely singled out. (An unfortunate feature of the word 'interpretation' is that it tends to obscure this: 'interpretation' is naturally taken to be equivalent to 'translation into our language', but of course that is *not* the sense of interpretation that has been defined.)

19. Here it is essential to use interpretations rather than models; for we want to hold that the relevant interpretation of the language of sets and classes has quantifiers ranging over all sets and classes, and these don't form a set or class.

20. The first example being that of Ayer and Ramsey.

## References

- Ayer, A. J. 1936. *Language, Truth, and Logic*. London: Gollancz.  
 Block, N. 1985. "Advertisement for a Semantics of Psychology." *Midwest Studies in Philosophy* 10: 615–678.

- Burge, T. 1979. "Individualism and the Mental." *Midwest Studies in Philosophy* 4: 73–121.
- Field, H. 1972. "Tarski's Theory of Truth." *Journal of Philosophy* 69: 347–375.
- Field, H. 1973. "Theory Change and the Indeterminacy of Reference." *Journal of Philosophy* 70: 462–481.
- Field, H. 1977. "Logic, Meaning, and Conceptual Role." *Journal of Philosophy* 74: 379–409.
- Field, H. 1978. "Mental Representation." *Erkenntnis* 13: 9–61.
- Field, Hartry. 1986. "The Deflationary Conception of Truth." In *Fact, Science, and Morality: Essays on A. J. Ayer's Language, Truth, and Logic*, edited by G. Macdonald and C. Wright. Oxford: Blackwell.
- Grover, D., Camp, J., and Belnap, N. 1975. "A Prosentential Theory of Truth." *Philosophical Studies* 27: 73–125.
- Horwich, P. 1982. "Three Forms of Realism." *Synthese* 51: 181–201.
- Kripke, S. 1975. "Outline of a Theory of Truth." *Journal of Philosophy* 72: 690–716.
- Leeds, S. 1978. "Theories of Reference and Truth." *Erkenntnis* 13: 111–129.
- Lewis, D. 1969. *Convention*. Cambridge: Harvard University Press.
- Lewis, D. 1975. "Languages and Language." In K. Gunderson, ed., *Minnesota Studies in the Philosophy of Science*, vol. 7. Minneapolis: University of Minnesota.
- Loar, B. 1981. *Mind and Meaning*. Cambridge: Cambridge University Press.
- Putnam, H. 1975. "The Meaning of 'Meaning'." In K. Gunderson, ed., *Minnesota Studies in the Philosophy of Science*, vol. 7. Minneapolis: University of Minnesota.
- Quine, W. 1970. *Philosophy of Logic*. Englewood Cliffs: Prentice-Hall.
- Ramsey, F. 1927. "Facts and Propositions." *Aristotelian Society*, suppl. vol. 7: 153–170.
- Soames, S. 1984. "What Is a Theory of Truth?" *Journal of Philosophy* 81: 411–429.
- Strawson, P. 1950. "Truth." *Aristotelian Society*, suppl. vol. 24: 129–156.



## The Prosentential Theory: Further Reflections on Locating Our Interest in Truth

Dorothy Grover

There are charges that deflationists have separated truth from the important issues.<sup>1</sup> It has also been charged that deflationary theories of truth are “philosophically more boring” than might be expected.<sup>2</sup> Such complaints will seem unequivocal to those who locate all that’s interesting about truth in a truth property. But arguing from the perspective of a deflationist, I will show why both charges are misconceived. In addressing the first charge, I will demonstrate, for a variety of contexts, the expressibility that a deflationary truth predicate provides. My goal will be to explain *how a* prosentential theorist has the option of subscribing to theses that place importance on truth, should he or she want to do that. This exploration will also serve to show that the *neutrality* of the prosentential theory with respect to the importance of truth is not a *denial* that truth is important. In addressing the second charge, I will remind readers of the challenge that deflationary theories pose for truth-property theories and of the advantage of not encumbering inquiry with the task of identifying a truth property.

### 1 The Prosentential Theory

I begin with a brief review of the prosentential account of the role of the truth predicate.<sup>3</sup>

I initially employed the concept of a prosentence (introduced by analogy with pronouns) to provide a reading of the anaphoric occurrences of bound propositional variables. Consider, for example, the following formula:

$\neg\exists p$  (Albert believes that  $p$  and  $p$ )

A reading in English will look like the following:

There is nothing such that Albert believes it is true, and it is true.

More colloquially, we might say this:

Nothing Albert believes is true.

'It is true' is used in providing a reading in English because it can accomplish the same kinds of logical connections that bound propositional variables accomplish, for each provides ways of making the logical connections we need for expressing generalizations. The basic claim of the prosentential theory is that 'it is true' and 'that is true' function as pro-sentences in English. Much as pronouns are used for generalizing with respect to nominal positions, so 'it is true' tends to be used for generalization with respect to sentence positions. 'That is true', on the other hand, tends to be used much as "pronouns of laziness" are used, but whereas a pronoun occupies a position that a noun occupies, 'that is true' occupies a sentential position. This means that in simple cases, 'that is true' stands in for, and anaphorically connects with, something else that has been said in the context. For example, 'that is true' has 'December 20 is the longest day' as its antecedent in the following dialogue:

*Mary* December 20 is the longest day.

*Tom* That is true, but only in the Southern Hemisphere.

Similarly, 'she' has the name 'Janet' as its antecedent in this context:

Because Janet likes to ski, she will not want to go hiking with us.

Just as 'she' picks up its referent from the name 'Janet', so 'that is true' will acquire its content from Mary's statement. Tom in effect is saying, but with anaphoric overtones, 'December 20 is the longest day, but only in the Southern Hemisphere'. This is partially what it is for 'that-is-true' to function as a prosentence. In the above dialogue, 'that is true' is used to say (with anaphoric overtones) something about December 20, and not to say something about a property of a sentence. In this lies an important difference between the prosentential account and those deflationary theories that represent 'true' as a metalinguistic predicate. The prosentential truth predicate typically keeps discourse at the level of the "object language,"<sup>4</sup> which means that if extralinguistic matters are being discussed, they remain the topic of discussion when 'true' is used. By con-

trast, other deflationary theories (e.g., Horwich 1990) represent 'true' as functioning at the level of the metalanguage.

Prosentences can be modified, as in, 'that was true', 'that might be true', and 'what is true?' For example:

*Tom* That was true last year, but this year the longest day is December 21.

In such simple cases, the modified prosentence can be said to stand in for a modified form of the antecedent, which in this case would be 'December 20 is the longest day', while the modified form would be 'December 20 was the longest day last year'. This prosentential picture serves to explain many pragmatic features of our truth talk, features highlighted in Strawson's 1950 work on truth [chap. 19—Ed.]. For example, in this context, use of 'that is true' allows Tom to implicitly acknowledge that it has already been mentioned that December 20 is the longest day. Anaphoric devices seem essential if we are to establish the connections needed for communication, for they facilitate the process whereby different speakers are able to talk about, and know they are talking about, the same things; they also provide logical connections needed for generalization.<sup>5</sup>

I will often appeal to the fact that generalizations have instances. Remember that 'it is true' tends to be used for generalization, as in the following:

I believe all that he said, but if you have reason to think some of it is not true, let us know immediately.<sup>6</sup>

I have found that in those cases where 'true' is used to express a generalization, the form of the prosentential reading is often more effectively conveyed (to those with a background of logic) through a paraphrase that uses bound propositional variables. In this case the propositional variable paraphrase could take the following form:

$\forall p$  (if he said that  $p$ , then I believe that  $p$ , but if you have reason to think that not  $p$ , then tell us immediately)

Its instances will include this:

Because he said that the nuclear reactor would be safe at the turn of the millenium, I believe it will be safe, but if you have reason to think the

nuclear reactor will not be safe at the turn of the millenium, let us know immediately.

I will continue to offer paraphrases using propositional quantifiers and variables.

Just as an account of the functioning of pronouns will be neutral with respect to many philosophical questions, so also the prosentential theory will be *neutral*. That is, the prosentential theory does not itself yield answers to general philosophical questions about meaning and communication, the status of science, realism, and so on.

The prosentential theory has been classified as a deflationary theory, a theory that denies there is (a need for) a truth property. This seems to follow from the fact that deflationary theories claim that the truth predicate has primarily a logical role. For if our use of the truth predicate is explained in some way in terms of anaphora, that is, in terms of 'true' as a prosentence-forming predicate, it can no longer be assumed *without argument* that 'true' must have a property-ascribing role. The prosentential theory, together with other deflationary proposals, poses a significant challenge for truth-property theorists, for it puts the onus on property theorists to show that the truth predicate ascribes a property.<sup>7</sup> The challenge has been taken seriously. Devitt (1984) and Field (1986), for example, have responded with the suggestion that if truth can be shown to have an explanatory role, then we need a truth property. Accordingly, they have offered suggestions for the explanatory role.

Another wrinkle also needs to be considered. I considered the possibility in Grover 1977 that a property-ascribing role might be "superimposed" on the anaphoric role. There are several reasons why I reject this possibility. First, I have not found an account of a truth property that is plausible, nor do I see a need for a property.<sup>8</sup> This is one reason why I have advanced, and continue to advance, the prosentential alternative. And except in the context of formal languages, I do not see an appropriate identification of the property's bearers. I have now also given reasons for thinking that the problem of inconsistency that seems to arise with Liar sentences is only introduced if a property-ascribing role is everywhere superimposed on the prosentential role (Grover 1977).<sup>9</sup> Since I see no reason to superimpose a property-ascribing role, and good reasons not to, I will continue to classify the prosentential theory as a

deflationary theory. The assumption that the truth predicate has only a logical role leaves the relative neutrality of the theory intact. The only exception to the neutrality thesis is that the prosentential theory denies that there is a substantive truth property.

## 2 Separation

Before beginning my defense that the prosentential theory does not separate truth from interesting and important philosophical issues, I must acknowledge that some aspects of the prosentential theory encourage philosophers to think that truth has been separated from all that is interesting and important. There is the fact that the prosentential theory, along with other deflationary theories, has been referred to as a theory of *truth*.<sup>10</sup> Described as a theory of *truth*, philosophers might reasonably expect that the prosentential theory provide answers to those questions that have been associated with other theories of truth: philosophical questions about meaning, belief, assertion, and the success of science, perhaps. However, as I have explained above, the prosentential theory provides only a theory of the truth predicate and is neutral on theories of meaning, belief, etc. Denying the need for a substantive truth property also encourages the view that truth has been separated from the important issues, especially among those property theorists who think that no truth property means no truth.

These features (its classification as a theory of *truth*, its neutrality, and the denial of a substantive truth property) certainly seem to suggest that the prosentential theory must separate truth from the important issues. How, then, can I claim that a prosentential theorist has the option of affirming the importance of truth in the event that he or she might want to?

My answer appeals to the fact that a prosententialist can accept that we have cognitive attitudes toward what-is-true and toward what-might-be-true, where 'true' is read as a prosentence-forming predicate. Indeed, my point will be that those concerned with the importance of truth would do well to use a prosentential predicate in formulating general questions and stating general theses in their approach to the "big" philosophical issues.



In my first paper on locating our interest in truth (Grover, forthcoming) I respond to Misak's (1998) charges of separation by showing that a prosentential theorist can accept claims like 'Our survival depends on having knowledge of what-is-true' and 'The goal of inquiry is to determine what-is-true'. Note that on a prosentential reading, 'what-is-true', just like 'Snow is white' and 'Electrons carry a negative charge', is ordinarily used to talk about extralinguistic things. There is only the difference that 'what-is-true' involves a generalization—a generalization like that in 'Our survival depends on having knowledge whether  $p$ , for many  $p$ '. The latter says that our survival depends on having knowledge: perhaps knowledge of the date of the longest day, the charge on electrons, the state of the ozone layer, and so on. Note that the instances of 'The goal of inquiry is to determine what-is-true' include 'The goal of inquiry is to determine whether December 20 is the longest day'. A prosentential theorist can also convey an interest in what-is-true (the truth) by asking questions: Could all human life be destroyed by a nuclear accident? Is the effect of pollution on the ozone layer irreversible? Does music increase intellectual performance? These are questions concerning how the world is, questions that are instances of the question 'What-is-true?'

### 3 What Is a Theory of Truth?

In this section I will show why I think a deflationary theorist can (consistently) explore the "big" issues, even to the extent of endorsing a theory of *truth*. But, first, if the prosentential theory is not a theory of *truth* and the correspondence theory is presumably not either (there being no truth property), where will we find a theory of truth?

In addressing Misak's concerns, I distinguished among *theories of the truth predicate*, *theories of what-is-true*, and *theories of theories of what-is-true*. The prosentential theory, together with theories that advance views concerning the logic of the truth predicate, are theories of the truth predicate. To the extent that science, creative endeavors, and value statements purport to tell us what-is-true, that is, tell us the way the world is, they are each theories of what-is-true. It would go too much against tradition to call theories of what-is-true (for example, scientific theories) theories of truth, but it would not be a mistake to do so.

Theories of theories of what-is-true are better candidates for being called theories of truth. Included are theories that tell us about the status of theories purporting to be theories of what-is-true. For example, they tell us whether theorists who claim to tell us about how the world is do indeed tell us how the world is. Does science tell us what-is-true? Should we follow Strawson (1985) in according legitimacy to the humanistic standpoint that we, as persons, occupy? Philosophy of science is one area of philosophy that yields theories of theories of what-is-true. Theories that address questions as to whether moral statements and literary statements tell us about the world also fall under this heading. So while theories of what-is-true and theories of theories of what-is-true address issues of truth, they do so in different ways. Theories of what-is-true are typically (though not always) concerned with determining how the extra-linguistic world is, while theories of theories of what-is-true are concerned with identifying the assumptions of inquiry itself as well as with assessing the methods used to determine what-is-true. The prosentential theory, in its attempt to describe the role of 'true', qualifies not only as a theory of the truth predicate but also as a theory of what-is-true; it is not, however, a theory of theories of what-is-true. Theories of assertion and belief, which may make significant appeals to truth, will usually qualify as theories of what-is-true, but not as theories of theories of what-is-true, except in so far as they are incorporated in theories of theories of what-is-true.

Note that the debates that arise among theorists of what-is-true and among theorists who assess the status of theories of what-is-true are much more complex than those arising among competing correspondence theorists. The latter focus on language-world connections or meaning, without concern for determining how the world is, that is, without trying to determine, for example, whether electrons have a charge, whether pollution affects the ozone layer, or whether any objects are basic in the physical universe. Of course, issues of language arise in debates among theorists of theories of what-is-true. But this is because issues of language and meaning are intertwined with issues in ontology and epistemology. I borrow a couple of questions from Kuhn (1970, 184) to illustrate the subtleties involved: "Did Einstein show that simultaneity was relative or did he alter the notion of simultaneity itself? Were

those who heard paradox in the phrase ‘relativity of simultaneity’ simply wrong?” While some of the hardest language questions may arise in our consideration of what-is-true, questions in the philosophy of language are not usually the primary focus of either theories of what-is-true or theories of theories of what-is-true. I assume that the primary focus is to determine whether  $p$  (for many  $p$ ); it is primarily our interest in extra-linguistic matters that motivate the task of assessing the subtleties of communication and our assumptions and methodology when determining whether  $p$ .

Truth-property theorists may be tempted to jump in and claim that theorists of theories of what-is-true are *really* identifying a truth property of sentences. But this is to miss the prosentential point that our interest in truth amounts to no more or less than our interest in knowing the way the world is. Any focus on sentences, or even utterances, is only a means to this end. Furthermore, I follow Tarski (1936) in thinking that it is only in the context of formal languages that extensions for ‘true’ and ‘false’ can be defined. While formal languages have many important uses and formal language analyses provide invaluable insight into the subtleties and logical structure of natural languages, I believe that they are not stand-alone languages of communication. This is because formal languages have been trimmed of essential features needed for communication.<sup>11</sup> So an extension for ‘true’ in a formal language has limited, though very important, interest.

Some deflationists, for example Horwich (1990) and Quine (1970, 10–13), assume that ‘true’ is a metalinguistic predicate and that ‘true’ has an extension. I accept this possibility for formal languages. For Quine, at least, it is the canonical language that matters, anyway. As I have argued elsewhere (Grover 1975, Grover 1990a) use of a metalinguistic predicate can be incorporated in the prosentential account. But note, in formal languages, instances of the T-schema, “‘ $p$ ’ is true iff  $p$ ,” will be relied on to bring talk of sentences back to talk about the way the world is. We need this connection, facilitated by a deflationary ‘true’, if formal languages are to be useful for more than just symbol-pushing maneuvers.

But let’s return to the present focus of our interest in truth. Can deflationists (those who deny there is a truth property) endorse a theory of truth? That is, can deflationists endorse a theory of theories of what-is-

true? Certainly they can. Quine's realist view of science (which is quite independent of his deflationary view of the truth predicate) qualifies as a theory of a theory of what-is-true. Consider, for example, the realist view expressed here:

But I also expressed, at the beginning, my unswerving belief in external things—people, nerve endings, sticks, stones. This I reaffirm. I believe also, if less firmly, in atoms and electrons and in classes. Now how is all this robust realism to be reconciled with the barren scene that I have just been depicting? The answer is naturalism: the recognition that it is within science itself, and not in some prior philosophy, that reality is to be identified and described. (Quine 1981, 21)

Whereas I see Quine as able to defend both a deflationary view of the truth predicate and a view of science according to which it describes the world, O'Leary-Hawthorne and Oppy (1997) think there is an inherent problem. They argue that Quine needs a supersense of truth. (They introduce the supersense of truth as follows: "The 'no super sense' theory claims that, if one has entertained the question whether *P* and come to a decision, there is no interesting further sorting procedure to be undertaken whether '*P*' is true, '*P*' correspondences to reality and so on" [O'Leary-Hawthorne and Oppy 1997, 180].<sup>12</sup>) The addition of a supersense of truth contradicts the spirit of Quine's deflationism. O'Leary-Hawthorne and Oppy make their criticism this way:

Quine assigns a thin role to the ordinary truth predicate, according to which it is merely a device for generalizing "along a dimension that cannot be swept out by a general term." He also holds there are discourses which do not "limn the truth." Given that we have the same need for generality in those other discourses too, there is considerable pressure here to introduce a supersense of "truth." (1997, 181)

I do not know why O'Leary-Hawthorne and Oppy think limning the truth calls for a supersense of truth. I can only think they do not recognize a deflationary 'true' in the phrase 'limn the truth'. Since no bibliographic reference is provided, I asked Peter Hylton for a suggestion. He has directed me to a passage in *Word and Object* that probably provides the basis of Hawthorne and Oppy's remark:

If we are limning the true and ultimate structure of reality, the canonical scheme for us is the austere scheme that knows no quotation but direct quotation and no propositional attitudes but only the physical constitution and behavior of organisms. (Quine 1960, 221)

On a deflationary reading, ‘truth’ (in ‘limn the truth’) and ‘true’ (in ‘limning the true’) would be used to express a generalization. If we are “limning the true and ultimate structure of reality,” we are providing illumination on the way the world is. For Quine, this is the task of science, for it is science that enlightens us as to whether  $p$ : science has the task of telling us whether electrons exist, and if they do, whether they have mass, and so on. And so I interpret Quine as recommending that his austere canonical language be the language of inquiry into what-is-true, the language of science.

The fact that Quine’s remarks can be interpreted as employing a deflationary truth predicate shows that he can consistently be a deflationist while at the same time endorsing a view according to which we should look to science to determine what-is-true. Because Quine can use a deflationary predicate to state his position that science determines whether a proposition is true, he remains consistent with the no-supersense position. So, though Quine has significant views about truth (he has a theory of theories of what-is-true), this is consistent with his deflationary view. Having said this, I am fully aware, as I mentioned earlier, that the task of determining whether  $p$  (for any  $p$ ) is very complex, raising in the long run, as it does, the deepest of issues in metaphysics, epistemology, and language.

#### 4 Linguistic Competence and the Concept of Truth

How might a prosententialist respond to criticism that claims that “the concept of truth” is important? This question arises because O’Leary-Hawthorne and Oppy use the phrase ‘the concept of truth’ throughout their commendable critique of deflationary theories. Consider how they identify a version of a minimalist view of truth:

One such candidate is that “is true” has a thin conceptual role: *mere possession of the concept of truth* contributes little or nothing to our understanding of the structure of reality and our relationship to it. The most radical version of this thesis will maintain that our grasp of other central cognitive concepts—belief, assertion, meaning, proposition, statement, translation, synonymy, fact, declarative sentence, negation, propositional connective, deep structure, logical form, semantics, etc.—is never explained in terms of the concept of truth; it will maintain further that deployment of the concept of truth does not, in and of itself, immediately commit us *a priori* to any interesting metaphysical theses. (1997, 174)

And then for the purpose of distinguishing further versions of minimalism, O'Leary-Hawthorne and Oppy raise the following questions.

How much does our understanding of various aspects of our conceptual scheme depend on our grasping the concept of truth? Does the making of judgments at all require possession of the concept of truth, as the Fregean idea of judgment—as advancement from thought to truth value—might suggest? Does a grasp of logical laws depend upon possession of the concept of truth? (1997, 178)

Philosophers who talk of a concept of truth assume concepts must be grasped if a property-ascribing predicate is to be properly applied. As deflationists deny that there is a truth property, they will also deny there is such a concept of truth. Because the property theorist's concept-of-truth talk is not easily adapted to the deflationist context, I need to find some other avenue of common ground if I am to provide a constructive response to O'Leary-Hawthorne and Oppy's remarks. They think that aspects of linguistic competence (a grasp of cognitive concepts, making judgments, a grasp of logical laws) require a grasp of the concept of truth.<sup>13</sup> For comparison, I will delineate two ways in which a prosententialist might appeal to an understanding of 'true' and 'what-is-true' in assessing linguistic competence.

We can certainly understand or fail to understand use of the truth predicate. For while most of us use the predicate appropriately, there could be people who do not have 'true' in their vocabulary. Young children come to mind, since just as there will be a stage at which a child cannot use pronouns satisfactorily, so she may not be able to use 'true' correctly. So my first suggestion for drawing a comparison will be that a speaker can be judged to have a certain level of linguistic competence when she has mastered our use of the word 'true', as well as use of its derivatives.

Connecting back to the specific remarks of Hawthorne and Oppy, we can note that such an understanding "contributes little or nothing to our understanding of the structure of reality." For just as our understanding of pronouns does not give us insight into the structure of reality, so also our understanding of how to use the word 'true', in itself, does not give us insight into the structure of reality. Nor will this feature (lack of insight) suffice for separating off a version of minimalism, because all deflationists will agree that an understanding of the use of 'true' does not lead to special insights. But deflationists will claim that the theoretical under-

standing that we offer of 'true' as having a logical role will help us sort through philosophical issues that have previously been encumbered by truth-property talk.

I remarked earlier that a prosententialist can use the truth predicate to talk about cognitive attitudes towards what-is-true and what-might-be-true. We can know what-is-true, wonder if we believe what-is-true, try to assert only what-is-true, be confused as to what-is-true, question whether *p*, for some *p*, and so on. We can also question whether a speaker can determine what is true on any occasion; that is, we can question whether a given speaker is able to determine whether *p*, for some *p*.

On occasions, some of us have trouble determining the difference between fantasy and reality, but there can be more serious failures. I have in mind cases where there is a complete failure to distinguish between what-is-true and what-is-false. This would mean that the speaker cannot tell whether snow is white, whether grass is green, whether she has a bank account, and so on. This seems to be the situation with people who have been referred to as *chatterboxes* or *blatherers* (Pinker 1994, 50–54).<sup>14</sup>

Chatterboxes have been described as “linguistic idiot savants—people with good language and bad cognition” (Pinker 1994). They can speak in syntactically well-formed sentences, string sentences together to make a story, and have impressive vocabularies, but they have limited cognitive contact with reality. There are reports of a fourteen-year-old who “chats on” about the trouble she has been having with the bank, the bank statements she has received, a joint account she has with her boyfriend, plans to go and talk with the bank, and so on. But she doesn’t have a bank account nor does she have a boyfriend. Indeed, chatterboxes are described as being “incompetent at ordinary tasks like tying shoes, telling left from right, adding two numbers, drawing a bicycle, suppressing their natural tendency to hug strangers” (Pinker 1994, 52). So though chatterboxes string sentences together in a relatively coherent way, their interactions with the world are seriously undeveloped. These cognitive shortcomings seem to show that chatterboxes have little or no *propositional knowledge*.

With respect to the issues to hand, it would seem possible that a chatterbox might correctly follow the syntax of 'true' and 'false' to the

extent that she can form well-formed sentences containing these words, perhaps as a parrot or robot might. But there is an important sense in which she hasn't really latched onto language. For if a chatterbox could not tell whether  $p$ , for any  $p$ , it would seem she could not have a goal of speaking the truth. This in turn would suggest that she could not be described as making assertions, nor described as knowing what counts as an answer to a question. Being able to determine, for at least some instances, whether snow is white, grass is green, etc., is clearly an essential ingredient of acquiring cognitive attitudes. (I assume that linguistic competence at the level of making assertions and asking questions involves the acquisition of other cognitive abilities.)

There is now a better match with the concerns of O'Leary-Hawthorne and Oppy. They think our grasp of cognitive concepts is to be explained in terms of the possession of the concept of truth—an ability to apply the truth predicate. Analogously, a prosententialist has the option of endorsing the thesis that a competent speaker must be able to tell whether  $p$ , for some  $p$ . That is, linguistic competence requires propositional knowledge, knowledge of some truths. O'Leary-Hawthorne and Oppy's implied claim that possession of the concept of truth contributes to our understanding of reality also has an analogue. A prosententialist has the option of endorsing the trivially true thesis that a speaker must be able to tell what-is-true in at least some instances if she is to "understand the structure of reality and our relationship to it."

It is interesting to note a contrast at this stage. In my first example I hypothesized that at a certain stage a child might have some communication abilities but not have either 'true' or 'false' in her vocabulary. If the child's communication skills are otherwise reasonably developed, it would be reasonable to say that she knows whether  $p$  for at least some  $p$ . In the case of chatterboxes, the reverse happens. A chatterbox may (in some limited sense) have 'true' and 'false' in her vocabulary but not be able to tell whether  $p$ , for any  $p$ . The former may be linguistically competent, while the latter is not.

Note a difference between O'Leary-Hawthorne and Oppy's property-ascribing account of linguistic competence and the one I am proposing. O'Leary-Hawthorne and Oppy seem to think that linguistic competence depends on our having grasped the concept of truth, which means (I am



supposing) that we should know when sentences or statements have the property of being true. But the example of the child who is acquiring language shows that a speaker might be reasonably linguistically competent yet not have 'true' in her vocabulary. The prosentential account opens up the alternative that linguistic competence requires that a speaker be able to determine whether  $p$ , for some  $p$ . That is, a linguistically competent speaker should be able to determine, in a "sufficient number" of cases, whether snow is white, whether grass is green, whether she has a bank account, and so on. There is no requirement that 'true' be included in the speaker's vocabulary or, in O'Leary-Hawthorne and Oppy's sense, that there be a grasp of the concept of truth.<sup>15</sup>

Once again, we have seen that a prosentential theorist can employ the truth predicate in talking about interesting and important issues. Not only can we describe cognitive attitudes in terms of attitudes toward what-is-true, what-might-be-true, etc. We can adopt, for example, the view that we are linguistically competent only if we have an ability to question and/or assert what-is-true.

## 5 Explanatory Role

Another charge that has often been made (in the context of the charge that deflationists separate truth from interesting issues) is to the effect that no appeal could be made by deflationists to truth in an explanation of the success of science.

O'Leary-Hawthorne and Oppy present this particular concern as a hypothetical scenario, as follows:

Suppose, for example, that science is converging on truth. Must it be simply illegitimate to explain the success of science in terms of this convergence, no matter what the context? ... Bald claims about the non-explanatory value of truth are normally accompanied by little more than a few examples. What we need is a theory of explanation which sets clear standards of explanatoriness and which provides some general account of why it is that truth cannot meet those standards. (1997, 182)

O'Leary-Hawthorne and Oppy assume that deflationists deny truth has an explanatory role in the hypothesized situation in which science converges. Certainly deflationists have denied that convergence in science would show an explanatory role for a truth property. I, for one, have

criticized Putnam's argument (1978, lectures II and III) that convergence in scientific theories is best explained by a correspondence account of truth and Field's attempts (1986) to discover an explanatory need for a correspondence truth property.

Initially, those who subscribed to the thesis that 'true' ascribed a property tended to make the following inference:

The word 'true' is used in (scientific) explanation.

So truth is a property with an explanatory role.

But such an appeal ignores the alternative of a deflationary reading of the word 'true'.<sup>16</sup> And so, initially, I responded by pointing out the possibility of a prosentential reading. Prompted by O'Leary-Hawthorne and Oppy's remarks, I now elaborate on the prosententialist position by showing that deflationists can certainly hold that truth can be important in explanations. Rather than construing explanations as appealing to a truth property, deflationists can construe explanations as appealing to what-is-true (the way the world is) and/or cognitive attitudes we have towards what-is-true, what-might-be-true, and so on.

Recall that prosentences function a bit like pronouns (when used anaphorically) and bound variables. One thing this means is that truth talk ordinarily proceeds at the level of the object language, rather than at the level of a metalanguage: truth talk is ordinarily (but not always) used to talk about how the extralinguistic world is. This feature of a prosentential account of 'true' will be most pertinent in my account of truth in explanations, just as it has been in the other cases considered.

Let us consider an example. Suppose, with respect to a certain subject matter (like cooking), that Sally's predictions are true, while Blake's are sometimes false. And suppose that the explanation of this difference is that Sally's relevant assumptions are true, while some of Blake's relevant assumptions are false. (It might be that Sally has more cooking experience, and so she is able to predict on the basis of knowledge, while Blake has little experience and must make wild guesses.<sup>17</sup>) To make the prosentential structure explicit, I rephrase so that the word 'true' is used in the prosentence 'it-is-true'.

1. Each relevant assumption Sally makes, when predicting what will enhance the flavor of a dish, is such that it-is-true. (Using propositional

variables and omitting a few details, we can write this as follows: for any  $p$ , Sally assumes that  $p$  only if  $p$ .)

2. At least one relevant assumption Blake makes, when predicting what will enhance the flavor of a dish, is such that it-is-false. (There is a  $p$ , Blake assumes that  $p$ , and not  $p$ .)

3. Each prediction Sally makes with respect to flavor is such that it-is-true, while at least one of Blake's predictions is such that it-is-false. ((For any  $p$ , Sally predicts that  $p$  only if  $p$ ) and (there is a  $q$ , Blake predicts that  $q$ , and not  $q$ ).<sup>18</sup>)

Because quantifiers bind the prosentences, we need to examine an instance of this explanation to see what is going on. Suppose that Sally and Blake have been asked to predict whether it makes a difference to the flavor of a dish if the garlic is sliced or minced. Then we have the following argument:

- a. Based on experience, Sally knows that sliced garlic enhances the flavor of soup, pasta, and fish dishes better than minced garlic does, and these are her assumptions.
- b. Making a wild guess, Blake assumes that it makes no difference to the flavor of dishes whether garlic is sliced or minced. (Blake wonders if minced garlic might have the edge, since recipes usually suggest using minced garlic, but then he also wonders if this suggestion is made in the interests of providing an easy recipe.)
- c. Sliced garlic enhances the flavor of soup, pasta, and fish dishes better than minced garlic does.
- d. Sally's prediction is that sliced garlic enhances the flavor of a bean casserole better than minced garlic will. Blake predicts that sliced garlic is no better than minced garlic.
- e. Sliced garlic enhances the flavor of a bean casserole better than minced garlic does.

The task was to explain why Sally's predictions are true and Blake's are false. The explanation is that Sally's assumptions are based on relevant experience and are true, while Blake's are not so based and some of his assumptions are false. But note that in (a) through (c) of the explanation 'true' is not used. Rather, we learn something about each of the following:

- The assumptions that Sally and Blake make
- The extralinguistic context in which Sally and Blake made their predictions: the enhanced flavor that results from using sliced garlic, rather than minced garlic, in soup, pasta, and fish dishes

- The cognitive basis on which Sally made her assumptions: her knowledge of the effects that sliced and minced cloves of garlic have on flavor<sup>19</sup>
- The cognitive basis on which Blake made his assumptions: his limited knowledge of cooking with garlic

I believe that causal, statistical, or (just) plausible, connections are implicitly appealed to also. These might concern constancy in the effects of sliced and minced garlic on the flavor of a range of dishes.

The list contains no use of a truth predicate because I replaced the bound occurrences of the prosentence 'it-is-true' in (1) through (3) with sentences not containing prosentences (for example, with 'sliced garlic enhances the flavor of soup better than minced garlic does'). So the prosentential reading has explanations like the above appealing to worldly details and cognitive attitudes toward the way the world is or might be. In contrast to property theorists' accounts, the prosentential account keeps the explanation in the object language, so to speak.<sup>20</sup>

Does absence of the use of the word 'true' mean that truth does not have a role in the explanation? Not quite, because we require that an explanation only appeal to what-is-true. In my example, (1), (2), (a), (b), and (c) should tell us what-is-true. So when we evaluate an explanation, we *do* pay attention to *truth*: we want to know what-is-true; we want to know what features of the world serve to explain the explanandum. It is in at least this sense that truth is important in explanations.

I doubt that these grounds suffice for saying truth has an *explanatory role*.<sup>21</sup> But even if we were to say that truth has an explanatory role in the sense just delineated, I would not yet have addressed the challenge that O'Leary-Hawthorne and Oppy posed. For the above would assign truth an explanatory role in *every* satisfactory explanation. O'Leary-Hawthorne and Oppy focused on a role that the truth property is said to have in only a select group of explanations.

The challenge was to explain why Sally's predictions were true and Blake's false. This is where a property theorist jumps in with the claim that the truth property explains why it is that Sally comes up with a true prediction and Blake does not. It is not just a question of it being true that Sally made certain assumptions; the assumptions themselves are true. But, of course, for a prosentential theorist, saying that Sally's assumptions are true is not to characterize her assumptions as having the property truth. In my example we have premises like this:

Sally assumed that sliced garlic enhances the flavor of soup better than minced garlic does, and it-is-true.

This says, but without the anaphoric overtones, something like the following:

Sally assumed that sliced garlic enhances the flavor of soup better than minced garlic does, and sliced garlic enhances the flavor of soup better than minced garlic does.

The prosentential reading presents the explanation as appealing not only to the fact that Sally made a certain assumption but also to features of the world (what-is-true). There is also an appeal to the circumstances under which Sally adopted the assumptions she makes.<sup>22</sup> Truth has a role in explanations not because we appeal to an explanatory truth property but because we appeal (among other things) to features of the world we consider relevant to the world's being as predicted. An appeal to a truth property would provide an irrelevant detour.

Some opponents to this prosentential account may wish to focus on the fact that Sally's assumptions did not come out of the blue but were arrived at through experience. It was not just by chance that she had the assumptions she had and that her assumptions were true. Sally had relevant knowledge. This suggests that there is an appeal in the explanation to connections between her cognitive attitudes and how things are in the world.<sup>23</sup> Perhaps there are, somewhere here, the language-world connections that correspondence theorists have tried to characterize. But the issues seem far from those of the correspondence theorist: the question as to whether Sally's assumptions are wild guesses, or evidentially based, is surely an issue for epistemologists; correspondence theorists have characteristically addressed issues of meaning.

I suggest that the proposed explanation of the success of science would come out similarly. (I will not address the issue of convergence, because I do not know what philosophers are trying to capture with this concept.) Let us suppose a simple case where the laws of the science and the observations of the scientists are all true. The explanation of the fact (if it is a fact) that the science yields true predictions will thus appeal to the fact that science has laws that tell us what-is-true and also to the fact that the reported observations of scientists tell us what-is-true. As far as I

can see, nothing would be gained by introducing an appeal to correspondence properties of the sentences that state the laws or describe the observations.

While I have not addressed many other questions and doubts that critics have expressed of deflationary theories, I hope that the above succeeds in presenting a slightly less sketchy account of how it is that a prosentential theorist has the wherewithal to take a position on the importance of truth. This in turn should also show that the approach of deflationists would help philosophers think more clearly about issues that most of us regard as central in philosophy.

### Notes

I thank Jerry Kapus for his insightful criticisms and suggestions on an earlier draft.

1. In a very interesting paper on deflationism and pragmatist truth, Misak (1998) has charged deflationists with separating truth from the interesting and important issues.
2. O'Leary-Hawthorne and Oppy 1997, 170.
3. Readers who are unfamiliar with the prosentential theory are referred to Grover 1992 for a better introduction. I also recommend Robert Brandom's account of the theory, together with his elaborations (1994, 301–333).
4. I put 'object language' in shudder quotes because, unless we are employing a formal language, the phrase is used metaphorically. I use 'metalinguistic language' similarly.
5. While several of my early papers (reprinted in Grover 1992) cover details of the prosentential thesis, there remains much work to be done in working out the syntactical details, as Kent Wilson (1990) has ably shown. It is a little reassuring to know that the syntax of pronouns has not yet been adequately explained either.
6. I use the hyphens to emphasize a prosentential reading of the truth predicate.
7. I do not mean to suggest that only property theorists must argue for their assumptions.
8. I must admit that a problem I have in responding to correspondence theorists in general is that I do not really know what would qualify as a correspondence theory. I think (as undoubtedly everyone does) that sequences of marks on paper and sequences of sounds do acquire properties as they come to be used for communication. It would be helpful to have more information about which among these provide the kind of property that correspondence theorists seek.

There have been suggestions for a correspondence property, but I have not found them satisfactory. In Grover 1990, I have presented a critique of Field's

suggestion. Field's theory has the same language-world connections for true and false sentences. I have always assumed that a correspondence theory would have different language-world connections for true and false sentences. Kirkham (1992) and David (1994) define correspondence truth in terms of 'obtains', but neither of them explains 'obtains'. This is a serious omission because 'obtains' serves as a replacement for 'true.' (My reviews of their books, Grover 1995 and Grover 1997, raise this difficulty.) The shortcomings of Russell's proposals are well known.

Deflationists are left with only the constructive approach: that of showing how well philosophers are served by a deflationary truth predicate.

9. But see Kapus 1993 for a critique of my claims. I am not persuaded by his argument that inconsistency threatens the prosentential theory because I reject his assumptions with respect to content. I leave my defense until such time as I have developed my ideas on language.

10. Though I myself have referred to the prosentential theory as a theory of truth, I now think that doing so was misleading. I return to this issue in the next section.

11. I am presently working on a project that addresses some of the differences between formal and natural languages.

12. Note that O'Leary-Hawthorne and Oppy's 'P' is used differently from my lower-case 'p'. Their 'P' is (almost always) a place holder in schemas. Instances of the schema are generated through substitution of sentences for 'P'.

13. I am assuming that O'Leary-Hawthorne and Oppy think the answer is 'Yes' to the questions 'Does making judgments at all require the possession of the concept of truth...?' and 'Does a grasp of logical laws depend on possession of the concept of truth?'

14. I thank my colleague Charles Chastain for having drawn my attention to this literature several years ago.

15. A dogged property theorist may respond with the charge that a speaker cannot know that snow is white, for example, unless she knows that 'white' *applies* to snow. What's the role of 'applies'? This question reintroduces the debate between deflationists and property theorists, but this time the focus will be on 'applies'.

16. Not everyone is guilty, for example, Devitt (1984). O'Leary-Hawthorne and Oppy also appreciate that something more is needed, since they attempt to establish a truth property.

17. I have space to consider just the one scenario, but there are variations on this example that could be entertained. For example, we might suppose that Sally and Blake have the same background of experience, and yet Sally's predictions are true and Blake's false. It might then be explained that Sally comes up with better hypotheses and why. Or it could be that Blake has a poor memory or that he pays little attention to the flavor of food.

18. Not having expertise in the subject of explanation, I can only surmise that there must be a significant connection between events appealed to in the expla-

nation and the event being explained. In the present case, Sally's prediction of enhanced flavor cannot be explained in terms of Sally's (true) beliefs that  $2 + 2 = 4$  and  $3 + 7 = 10$ . An explanation must presumably be based on assumptions that make it reasonable that she arrived at the predictions she does. I have used 'relevant' to cover whatever features must be included.

19. I assume that Sally has relevant knowledge. If she had only been able to make wild guesses, then we might say that there is no explanation for her success beyond being lucky. I also assume that the effect of garlic on bean dishes is somewhat similar to its effect on other dishes.

20. This is how I construe this kind of explanation. Of course, there are explanations of linguistic matters where one will want to appeal to "meta-linguistic" properties.

21. Again, not having expertise in the area of explanation, I am inclined to accept the suggestion that properties are the only candidates for an explanatory role.

22. By contrast, for the property theorist, there is an appeal in the explanation to the fact that Sally's assumption is characterized as being true; truth, as a property, performs some explanatory work.

23. It is difficult for me to develop a stronger case for the opposition because, as I have mentioned in an earlier note, I am not quite sure how the correspondence theory is supposed to go. Because I do, of course, think that there are connections between the extralinguistic world and our acquiring language and cognitive attitudes, I cannot try to forestall correspondence theorists by claiming that language-world connections have *no* place in explanations. For example, perhaps a version of the causal theory of perception will feature in some explanations of the predictions of Sally and Blake. Whichever way this might come out, my guess is that the situation is much more complex than any correspondence theorist has so far represented it. As Jerry Kapus has said in a comment on this section of the paper, there are matters here that all of us have to address.

## References

- Brandom, Robert B. 1994. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge: Harvard Univ. Press
- David, Marian. 1994. *Correspondence and Disquotation: An Essay on the Nature of Truth*. Oxford: Oxford Univ. Press.
- Devitt, Michael. 1984. *Realism and Truth*. Princeton: Princeton Univ. Press.
- Field, Harry. 1986. "The Deflationary Conception of Truth." In *Fact, Science, and Morality: Essays on A. J. Ayer's Language, Truth, and Logic*, ed. Graham Macdonald and Crispin Wright, pp. 83–107. Oxford: Blackwell, 1986.
- Grover, Dorothy. 1977. "Inheritors and Paradox." *Journal of Philosophy* 74: 590–604. Reprinted in Grover 1992.
- Grover, Dorothy. 1990a. "On Two Deflationary Truth Theories." In *Truth or Consequences*, ed. Michael Dunn and Anil Gupta, pp. 1–17. Dordrecht: Kluwer. Reprinted in Grover 1992.



- Grover, Dorothy. 1990b. "Truth and Language-World Connections." *Journal of Philosophy* 87: 671–687. Reprinted in Grover 1992.
- Grover, Dorothy. 1992. *A Prosentential Theory of Truth*. Princeton: Princeton Univ. Press.
- Grover, Dorothy. 1995. Review of *Theories of Truth: A Critical Introduction*, by Richard Kirkham. *Philosophy and Phenomenological Research* 55: 3.
- Grover, Dorothy. 1997. Review of *Correspondence and Disquotation*, by Marian David. *Journal of Symbolic Logic* 62: 326–328.
- Grover, Dorothy. Forthcoming. "On Locating Our Interest in Truth." In *What Is Truth?* ed. Richard Schantz. Berlin and New York: de Gruyter.
- Grover, Dorothy, J. L. Camp, and N. D. Belnap. 1975. "A Prosentential Theory of Truth." *Philosophical Studies* 27: 73–124. Reprinted in Grover 1992.
- Horwich, Paul. 1990. *Truth*. Oxford: Blackwell.
- Kapus, Jerry. 1991. "The Liar and the Prosentential Theory of Truth." *Logique et Analyse* 34: 283–291.
- Kirkham, Richard L. 1992. *Theories of Truth: A Critical Introduction*. Cambridge: MIT Press.
- Kuhn, Thomas S. 1970. *The Structure of Scientific Revolutions*. Chicago: Univ. of Chicago Press.
- Misak, Cheryl. 1998. "Deflating Truth: Pragmatism vs. Minimalism." *Monist* 81, no. 3: 407–425.
- O'Leary-Hawthorne, John, and Graham Oppy. 1997. "Minimalism and Truth." *Noûs* 31, no. 2: 170–196.
- Pinker, Steven. 1994. *The Language Instinct*. New York: William Morrow.
- Putnam, Hilary. 1978. *Meaning and the Moral Sciences*. London: Routledge and Kegan Paul.
- Quine, W. V. 1960. *Word and Object*. Cambridge: MIT Press.
- Quine, W. V. 1970. *The Philosophy of Logic*. New York: Harper and Row.
- Quine, W. V. 1981. *Theories and Things*. Cambridge: Harvard Univ. Press.
- Strawson, P. F. 1950. "Truth." *Proceedings of the Aristotelian Society*, suppl. vol. 24: 129–156. Reprinted as chap. 19 of this volume.
- Strawson, P. F. 1985. *Skepticism and Naturalism: Some Varieties*. New York: Columbia Univ. Press.
- Tarski, Alfred. 1936. "The Concept of Truth in Formalized Languages." In his *Logic, Semantics, Metamathematics: Papers from 1923–1938*, trans. J. H. Woodger, pp. 152–278. Oxford: Clarendon Press, 1956.
- Wilson, Kent. 1990. "Some Reflections on the Prosentential Theory of Truth," In *Truth or Consequences: Essays in Honor of Nuel Belnap*, ed. J. Michael Dunn and Anil Gupta. Dordrecht, Netherlands: Kluwer Academic Publishers.

---

## A Critique of Deflationism

Anil Gupta

### I

Throughout much of this century there have been two types of philosophical debates over the concept of truth. In the first, *substantive*, type of debate we find rival theories of truth put forward that seem to have, and whose proponents have taken them to have, significant metaphysical and epistemological implications. An early example of this type is the debate in the early 1900s between the British Idealists (F. H. Bradley and his followers) and the Logical Atomists (Bertrand Russell and his followers). The Idealists defended a coherence theory of truth, whereas the Atomists argued for a correspondence theory. This dispute over the theory of truth was not, and was not taken by the participants to be, a local disagreement. It was integral to the larger metaphysical debate between the two sides over monism and pluralism and over idealism and realism. A recent example of the substantive type is the debate between the realist and the anti-realist found in Michael Dummett's writings. The crux of the debate here is what notion of truth is admissible. Dummett's anti-realist argues for a notion of truth that is constrained by evidence, while the realist defends the admissibility of a radically non-epistemic notion.

In the first type of debate, then, we find theses put forward and defended that have (or at least seem to have) substantial philosophical implications. Debates of this type presuppose that truth has a substantial role to play in philosophical inquiry. In the debates of the second, *metaphysical*, type the presupposition is called into question. An early example of this type is the debate over the claim, made by some Logical Positivists, that truth is a metaphysical concept and hence ought to be banished from all rigorous and scientific thought. A decisive contribution

to this debate was made by Alfred Tarski, who gave a definition of truth (for certain languages) that was adequate by the Positivists' own strictures. Tarski's definition used only terms that the Positivists found legitimate, and it defined a notion that was provably coextensive with truth. Tarski's work was widely viewed as establishing the legitimacy and the usefulness of truth in philosophical inquiry. One result of its influence was a shift away from a syntactical conception of language and towards a semantical one.<sup>1</sup>

Ironically, Tarski's work, while refuting one sort of skepticism about the usefulness of truth, provided a basis for a different, more compelling, kind of skepticism. This new kind of skepticism, *deflationism*, maintains that truth is a simple and clear concept and has no substantial role to play in philosophy. Substantive debates over truth, according to deflationism, are in error not because they work with a notion that is metaphysically loaded (and hence corrupt), but because they work with a notion that is metaphysically lightweight. Deflationism has provoked a large debate among philosophers—a debate that provides a contemporary instance of the second, metaphilosophical, type of debate distinguished above.

A deflationary view typically consists of two parts: (i) a description of the meaning and function of 'true' and (ii) a derivation from that description of deflationary consequences concerning truth. As an example of (i), consider the following passage from Michael Williams: It contains a popular account of the meaning and function of 'true.' (In the next section I shall explain and discuss the account in detail.)

[W]hen we have pointed to certain formal features of the truth-predicate (notably its 'disquotational' feature) and explained why it is useful to have a predicate like this (e.g. as a device for asserting infinite conjunctions), we have said just about everything there is to be said about truth.<sup>2</sup>

Examples of (ii) can be found in sections III and IV below. The following extracts illustrate the sorts of deflationary consequences that are often drawn. The first extract is from Richard Rorty and the remaining two are from Scott Soames and Paul Horwich.<sup>3</sup>

[T]ruth is not the sort of thing one should expect to have a philosophically interesting theory about.<sup>4</sup>

What does seem right about Tarski's approach is its deflationist character.... Truth is a useful notion, but it is not the key to what there is, or to how we represent the world to ourselves through language.<sup>5</sup>

[Truth is not] a deep and vital element of philosophical theory. . . . [T]he realism/anti-realism issue (together with various related issues in the philosophy of science) have nothing at all to do with truth.<sup>6</sup>

In short, deflationism holds that once we understand the meaning and function of 'true'—and this understanding, according to deflationism, is not hard to achieve—we shall see that truth has no substantial role to play in philosophy. Many contemporary philosophers find the deflationary account of 'true' attractive and plausible, and they have accepted (sometimes enthusiastically, sometimes regretfully) its negative verdict on the role of the concept of truth in philosophy.

I want to oppose deflationary attitudes in philosophy. The main problem with deflationism, in my view, lies in the descriptive account it gives of 'true.' The deflationary account makes (and, to sustain its conclusions, needs to make) some very strong claims about the meaning of 'true'—claims that on examination prove to be highly problematic. The account appears plausible, I think, only because we read its claims in a weaker way. But the weaker readings do not, I believe, yield any deflationary conclusions.

The argument I shall develop against deflationism, then, is this. The deflationary description of 'true,' when it is taken in the strong and intended way, motivates the deflationary conclusions, but is highly problematic. On the other hand, when it is taken in the weaker way, the description is correct enough, but does not yield the deflationary conclusions. I shall substantiate this by considering deflationary arguments on two issues: the possibility of a physicalistic theory of truth (section III), and truth and meaning (section IV). Deflationists take the concept of truth to be transparent, one capable of a complete and simple philosophical analysis. Towards the end of the paper (section V) I shall point out some reasons to think that truth is a highly puzzling notion, one that defies all our attempts at its analysis.

## II

Let us consider the *disquotational* account of the meaning of 'true,' which we encountered briefly in the extract from Williams.<sup>7</sup> Its original source is the following well-known passage from W. V. Quine's *Philosophy of Logic*.<sup>8</sup>

By calling the sentence ['snow is white'] true, we call snow white. The truth predicate is a device of disquotation. We may affirm the single sentence by just uttering it, unaided by quotation or by the truth predicate; but if we want to affirm some infinite lot of sentences that we can demarcate only by talking about the sentences, then the truth predicate has its use. We need it to restore the effect of objective reference when for the sake of some generalization we have resorted to semantic ascent.<sup>9</sup>

Stephen Leeds provides, in the following extract, a useful elaboration of the disquotational account.

It is not surprising that we should have use for a predicate *P* with the property that “‘\_\_\_’ is *P*” and “\_\_\_” are always interdeducible. For we frequently find ourselves in a position to assert each sentence in a certain infinite set *z* (e.g., when all the members of *z* share a common form); lacking the means to formulate infinite conjunctions, we find it convenient to have a single sentence which is warranted precisely when each member of *z* is warranted. A predicate *P* with the property described allows us to construct such a sentence:  $(x)(x \in z \rightarrow P(x))$ . Truth is thus a notion that we might reasonably want to have on hand, for expressing semantic ascent and descent, infinite conjunction and disjunction. And given that we want such a notion, it is not difficult to explain how it is that we have been able to invent one.<sup>10</sup>

The core thought here is that the function of the truth predicate is to serve certain expressive purposes, namely, that of expressing certain infinite conjunctions and disjunctions. The truth predicate serves these functions in virtue of its disquotational character; i.e., in virtue of the fact that it undoes the effect of quotation marks.<sup>11</sup> For example, the role of ‘true’ in

(1) ‘snow is white’ is true

is to cancel the quotation marks: (1) says no more nor less than the sentence

snow is white.

We shall get clearer on the disquotational theory if we consider a situation in which, as Quine puts it, “we want to affirm some infinite lot of sentences.” Suppose we wish to affirm all sentences of the form

\_\_\_ & snow is white [= *A*, say].

That is, we want to affirm the conjunction of all sentences obtained by filling the blank in *A* with sentences of English:

- (2) [Sky is blue & snow is white] & [Chicago is blue & snow is white]  
& ...

We lack explicit and direct means of formulating the infinite conjunction, but the truth predicate, according to Quine and Leeds, provides us with an indirect means. Observe that we cannot generalize on the '\_\_\_\_' position in *A* using ordinary first-order variables. We cannot say, for example, For all *x*: *x* & snow is white.

For the variable '*x*' is pronominal and occupies name positions; it cannot meaningfully be put in sentence positions. The way the truth predicate helps here, according to the disquotational account, is this. The disquotational feature of truth makes (2) equivalent to

- (3) ['Sky is blue' is true & snow is white] & ['Chicago is blue' is true  
& snow is white] & ...

But the position '\_\_\_\_' in  
\_\_\_\_ is true & snow is white

is nominal and can be quantified using the pronominal variable '*x*'. We can say,

- (4) For all sentences *x*: [*x* is true & snow is white].

But (4) is equivalent to (3) and, consequently, in virtue of disquotation, to (2). The truth predicate thus provides us with a means of expressing the infinite conjunction (2). Truth is, on the disquotational account, essentially a logical device. It enables us to generalize over sentence positions *while using pronominal variables such as 'x'* and, thus, endows us with additional expressive power.

It will be useful to separate out four component ideas of the disquotational theory.

*The Disquotation Thesis* The truth predicate is a device of disquotation.  
*The Infinite Conjunction Thesis* The truth predicate enables us to express certain infinite conjunctions and disjunctions; (4), for instance, expresses (2) and (3).<sup>12</sup>

*The Generalization Thesis* The truth predicate provides a means for generalizing over sentence positions even when the variables are pronominal.

*The Connection Thesis* The truth predicate serves its expressive functions in virtue of its disquotation feature.<sup>13</sup>

The first two of these theses contain important ambiguities. Let us demarcate a little the sense in which the deflationists understand these theses (and need to understand them).

Let us call instances of the form

(T) ‘\_\_\_\_’ is true if and only if \_\_\_\_

*T-biconditionals*.<sup>14</sup> Then, the Disquotation Thesis is understood by the deflationists as saying not just that the T-biconditionals are true, nor just that they are necessarily true.<sup>15</sup> The claim is rather that the T-biconditionals issue from our very understanding of ‘true,’ that they explain (at least partially) the meaning of ‘true.’<sup>16</sup> This way of reading the Disquotation Thesis is not always explicit in the writings of the deflationists. But, as we shall see, it is required by key deflationary arguments. Furthermore, some authors *are* explicit on the point. Horwich has stated that our understanding of ‘true’ *consists* in our “disposition to accept, without evidence, any instantiation of the schema [(T)].” And he speaks of the T-biconditionals as constituting a *definition* of ‘true.’<sup>17</sup> Even philosophers opposed to deflationism have often been attracted to this reading of the Disquotation Thesis.<sup>18</sup> Hartry Field’s influential paper “Tarski’s Theory of Truth” argues for a view as far removed from deflationism as any. Yet it contains a description of ‘true’ that would fit comfortably in any deflationary text:<sup>19</sup>

[L]et’s note one obvious fact about how the word ‘true’ is standardly learned: we learn to apply it to utterances of our own language first, and when we later learn to apply it to other languages it is by conceiving the utterances of another language more or less on the model of utterances of our own language. The obvious model of the first stage of this process is that we learn to accept instances of the schema

(T) *X* is true if and only if *p*.

where ‘*X*’ is replaced by a quotation-mark name of an English sentence *S* and ‘*p*’ is replaced by *S*.<sup>20</sup>

In summary, we shall understand the Disquotation Thesis as stating that disquotation provides an analysis of ‘true,’ that it explains (at least partially) what the word means and what our understanding of it consists in. The thesis should be sharply distinguished from weaker ideas such as that the T-biconditionals are necessarily true.

The Infinite Conjunction Thesis separates out for consideration the claim, often made by the deflationists (and sometimes by the non-deflationists also), that the truth predicate is a device for expressing certain infinite conjunctions and disjunctions. The thesis is ambiguous because of an ambiguity in 'express.' Is the thesis to be read so that it yields only that (4) and (2) are *materially* equivalent? Or that they are necessarily equivalent? Or that they have the same sense? Or something yet different? The deflationists have not been explicit on the point. We shall see, however, that the use they make of the Infinite Conjunction Thesis requires that 'express' be read in a strong way.

One argument of the deflationists—that for the Connection Thesis—requires much too strong a reading of the Infinite Conjunction Thesis. According to the Connection Thesis the truth predicate *needs* to be disquotational if it is to serve its expressive functions—in particular, its function of expressing certain generalizations. The argument for the thesis was implicit in our exposition above: The function of (4) is to express (2). But this is possible only if (2) and (3) are equivalent. Here is where disquotation comes in. It is needed to ensure that the equivalence of (2) and (3) holds. Hence, disquotation is needed to ensure that truth plays its desired role in generalizations such as (4). The role of the Infinite Conjunction Thesis in the argument is to show that (2) and (3) need to be equivalent, if the truth predicate is to play its expressive role. But this motivates the need for a disquotational truth predicate only if the equivalence of (2) and (3) is required to be something like sameness of sense. Anything weaker will yield the need, not for disquotational truth, but for something weaker. If, for example, the role of truth in (4) requires only that (2) and (3) be necessarily equivalent, then the argument will yield only that the T-biconditionals must be necessarily true if 'true' is to serve its role. It will not yield the Disquotation Thesis.

In the strong sense needed for the Connection Thesis, the Infinite Conjunction Thesis is plainly false. A universal statement (e.g., (4)) does not have the same sense as the conjunction of its instances (e.g., (3)). The two typically do not even imply the same things; they are equivalent only in a much weaker sense.<sup>21</sup> I think that the proponents of the disquotational theory have gone astray because they have ignored the difference between wanting to affirm a generalization and wanting to affirm each



of its instances. Quine writes in the passage quoted above, “if we want to affirm some infinite lot of sentences that we can demarcate only by talking about the sentences, then the truth predicate has its use.” In the situation envisaged by Quine, where we can demarcate some infinite lot of sentences only by talking about them, what we typically want to do is affirm a generalization—and the truth predicate enables us to do this. We can, for example, generalize on the ‘\_\_\_\_\_’ position in

\_\_\_\_\_ & snow is white

with the aid of the truth predicate, as we saw above. But this is not the same as affirming “some infinite lot of sentences,” which requires an infinitary conjunction. It is because two distinct things (which, to repeat, are affirming the universal and affirming all the instances) are confused that the infinitary conjunction seems to be strongly equivalent to the generalization, and leads in turn to the Disquotation Thesis. [I think the same confusion is going on in Leeds’ claim in the passage quoted above that “we frequently find ourselves in a position to assert each sentence in a certain infinite set  $z$  (e.g., when all the members of  $z$  share a common form).”]

The Connection Thesis, then, rests on a confusion. This, I think, is a blemish on deflationism: It means that the deflationary accounts of the meaning and the function of ‘true’ are not connected in the neat way that the deflationists supposed. But this does not damage deflationism in a material way. For the arguments for the deflationary attitude towards the role of truth in philosophy rest not on the Connection Thesis but on the Disquotation and the Infinite Conjunction Theses. Let us now examine some of these arguments.

### III

One question that philosophers have debated in recent years is whether truth is amenable to a physicalistic reduction—in other words, whether truth is a complex physical property. Two compelling philosophical pictures, when combined, suggest a positive answer: (i) the correspondence theory of truth and (ii) a physicalistic ontology. The former suggests that underlying truth there is a systematic relation between words and the

world; the latter suggests that this relation can be understood in physical terms. The combination of the two pictures, in fact, makes each a little more attractive. Traditional correspondence theories are confronted with the embarrassment that they have had little to say (beyond such vacuous claims as 'snow' refers to snow) about the relation between words and the world. Physicalistic ontology is useful here: It provides a framework in which a substantial account of the relation might be spelled out. Physicalistic ontology has faced, on the other hand, the problem of giving an account of psychological and semantic properties. A correspondence theory helps here: It provides a scheme for making sense of at least one semantic property. In short, the idea that truth is a complex physical property makes the two philosophical pictures a little more coherent and attractive.

The deflationist position on the question is, as one would expect, that truth is not amenable to a physicalistic reduction, that to suppose otherwise is to misunderstand the meaning and function of 'true.' We shall examine the deflationary arguments for this claim after we have briefly reviewed the debate within which the arguments arose.

Hartry Field initiated the debate by arguing (in his paper, "Tarski's Theory of Truth") that truth is amenable to a physicalistic reduction. Field argued that just as the usefulness of the concept of valence in chemistry is a reason to expect a physicalistic reduction for it, so with truth: The usefulness of the notion of truth is a reason to think that it has a physicalistic reduction. Stephen Leeds pointed out a problem with this argument.<sup>22</sup> What provides us with a reason to expect a reduction of "valence" is that it is a *causal-explanatory* notion, as is shown by its role in the *law* of valences. Mere usefulness does not establish the requisite analogy of truth with valence. What must be shown is that there are laws of truth analogous to the law of valences. (Leeds went on to suggest that the utility of truth can be explained by seeing it as a device for expressing infinite conjunctions and disjunctions.) Hilary Putnam took up Leeds' challenge.<sup>23</sup> He argued that, like valence, truth does play a causal-explanatory role. He offered several generalizations as examples of causal-explanatory laws involving truth—generalizations such as the following:

(5) The laws of a mature science are typically approximately true.

- (6) True beliefs about how to attain our goals tend to facilitate success in achieving them.
- (7) Beliefs formed as a result of our methods of inquiry tend to be true.

The first law, Putnam suggested, helps explain the success of the mature sciences; the last two *our* success in attaining our goals.<sup>24</sup> The deflationists responded that Putnam's examples do not pose a difficulty for them; the examples, they argued, can be explained within their framework. Their arguments seem to have been widely accepted and have contributed to the prevalent skepticism of the possibility of a physicalistic reduction of semantic concepts.

Let us consider how the deflationary arguments go for one of Putnam's examples. (The others are treated in parallel way.) Here is how Williams responds to (6).

I see no reason to think of [(6)] as a law. . . . If I want a cold drink and believe that the refrigerator, rather than the oven, is the place to look, I will increase the likelihood of getting what I want. This is because cold drinks are more likely to be found in the refrigerator than in the oven. To say that my having true beliefs makes it more likely that I will attain my goals is just a compact way of pointing to the indefinite number of mundane facts of this sort. It involves nothing so arcane as a physical correspondence theory of truth.<sup>25</sup>

Williams argues here that (6) is not a law, since it is "just a compact way of pointing to the indefinite number of mundane facts" of the sort he cites. Let  $A_1, A_2, A_3, \dots$  be these mundane facts. Williams' argument rests on the idea that (6) expresses—in some sense of "expresses"—the infinite conjunction

- (8)  $A_1 \ \& \ A_2 \ \& \ A_3 \ \dots$

It is plain that Williams' argument does not work if "express" is understood in an extensional way; i.e., if we suppose only that (6) is materially equivalent to (8). Nor does the argument work if we take "express" in an intentional way; i.e., if we suppose only that (6) is necessarily equivalent to (8). For, of two sentences that are necessarily equivalent, one can be a law and the other not. Here is an example:

- (9) Cicero is Tully.

- (10) No chemical reaction will produce caustic soda from saltpeter and sulfuric acid.

Both these statements are necessary truths and, hence, are necessarily equivalent.<sup>26</sup> The second states a law, but not the first. Only when the equivalence between two sentences is very strong can we infer the nomological character of one from the nomological character of the other. Williams' argument presupposes therefore a strong reading of the Infinite Conjunction Thesis.

Horwich responds to Putnam in a different way. He does not deny that (6) is a law. He argues instead that (6) is sufficiently explained by the T-biconditionals and, hence, that we do not need a substantial correspondence theory of truth to explain it. Horwich writes:

[I]t is clear, in general, how true beliefs contribute to practical success. Nothing beyond the minimal theory [which consists essentially of the T-biconditionals] is called for to explain this phenomenon.<sup>27</sup>

The way the T-biconditionals explain (6), according to Horwich, is this. Suppose that

- (11) Bill believes that he will get a beer if he nods  
and that

- (12) Bill wants a beer.

Sentences (11) and (12) explain Bill's nod. The truth of Bill's belief yields, in virtue of the T-biconditionals, that

If Bill nods, he will get a beer.

Hence, we obtain the conclusion that Bill will get a beer and, consequently, that his want will be fulfilled. Other examples of beliefs and desires may require a more complex explanation, but, as the above example illustrates, none will need a substantial theory of truth.<sup>28</sup>

This argument needs the support of the Disquotation and the Infinite Conjunction Theses to work. The argument invites two challenges. First, it may be argued that even if the T-biconditionals explain (6), the need for a substantial theory of truth remains. It may be that a substantial theory of truth will provide a deeper explanation of the T-biconditionals and, consequently, of (6). Second, it may be argued that what Horwich proposes is an explanation only of the instances of (6), not of (6) itself.

The Disquotation Thesis provides a response to the first challenge: Since the T-biconditionals are definitional of truth, the response goes, they are not open to a deeper explanation;<sup>29</sup> the substantial theory of truth has no work to do. Observe that this response will not work on the weaker readings of the Disquotation Thesis. It will not work, for example, if all we have available is the thesis that the T-biconditionals are necessary truths. For, necessary truths *can* sometimes be given a deeper explanation. Sentence (10) expresses a necessary truth, yet chemistry provides a deep explanation of why it holds.

The Infinite Conjunction Thesis provides a response to the second challenge: Since the T-biconditionals explain all the instances of (6), they explain their infinite conjunction. But by the Infinite Conjunction Thesis, (6) expresses this infinite conjunction. Hence, the T-biconditionals explain (6) also. Again, observe that this response requires a strong reading of 'express.' The necessary equivalence of (6) with an infinite conjunction is insufficient ground for it. For, to explain one of two sentences that are necessarily equivalent is not thereby to explain the other (see examples (9) and (10)).

We may conclude, then, that the deflationary arguments against a substantial theory of truth need the support of the Disquotation Thesis and the Infinite Conjunction Thesis read in a strong way; weaker versions of the theses are insufficient. This is a major weakness in the arguments, for on the strong reading the Infinite Conjunction Thesis is false. Williams' argument requires the generalization (6) to be equivalent to the infinite conjunction (8) in a sense strong enough to guarantee sameness of nomological character. But the two plainly are not equivalent in such a strong sense. The conjuncts of (8) are particular in character. So, (8) itself is particular in character. But this is not true of (6), which is general. Further, (6) gives us information about counterfactual situations that lie beyond the infinite conjunction (8).

Horwich's argument presupposes that a generalization is equivalent to the conjunction of its instances in a sense strong enough to guarantee that an explanation of one is an explanation of the other. But, as the following example shows, this is not true. We can explain each instance of the generalization "everyone on the boat died" by providing a separate explanation for the death of each person on the boat: Jack died of a

heart attack; Mohini drowned; etc. But these separate explanations do not necessarily explain the generalization. The generalization may in fact have no explanation at all—it may be true accidentally. Or it may have an altogether different explanation, such as that the boat capsized. In any case, an explanation of the instances is not necessarily an explanation of the generalization.<sup>30</sup>

We can accept the Infinite Conjunction Thesis when ‘express’ is understood as implying only material equivalence. We can even suppress several doubts and grant the thesis when ‘express’ is understood as implying necessary equivalence. But the thesis is false when ‘express’ is understood in the strong way needed in Williams’ and Horwich’s arguments.<sup>31</sup> I think the cause of error here is the same oversight that we found in Quine’s passage in section II: a neglect of the distinction between “affirming the universal” and “affirming all the instances.” Once the distinction is neglected it becomes easy to read the Infinite Conjunction Thesis in a strong way. Once the distinction is marked, the strong readings are seen to be plainly false.

In conclusion: The deflationary arguments against a substantial theory of truth presuppose an unacceptably strong reading of the Infinite Conjunction Thesis. I myself see nothing in the meaning and function of ‘true’ to rule out the possibility or the usefulness of a substantial theory of truth.<sup>32</sup>

#### IV

The theory of meaning is another area in which the deflationists deny truth a substantial role. Two paradigms dominate philosophical studies of meaning. One paradigm seeks to understand meaning in terms of language-*world* relations. On this paradigm the concept of truth plays a central role in an account of meaning. Indeed, on many theories within this paradigm, meaning (of a sentence) is identified with truth conditions. The other paradigm seeks to understand meaning in term of language-*user* relations. On this paradigm language-world relations are not so central in an account of meaning. What is central is the *use* to which sentences are put. The debate between the two paradigms is large and of large significance. The deflationist contribution to the debate is the argu-

ment that the meaning and function of 'true' rule out a truth-conditional account of the meanings of sentences.

An early formulation of the argument occurs in Michael Dummett's paper "Truth":

[I]n order that someone should gain from the explanation that *P* is true in such-and-such circumstances an understanding of the sense of *P*, he must already know what it means to say of *P* that it is true. If when he enquires into this he is told that the only explanation is that to say that *P* is true is the same as to assert *P*, it will follow that in order to understand what is meant by saying that *P* is true, he must already know the sense of asserting *P*, which was precisely what was supposed to be being explained to him.<sup>33</sup>

Dummett goes on to write, in the concluding paragraph of his paper, that

[F]or most ordinary contexts the account of these words ['true' and 'false'] embodied in the laws 'It is true that *p* if and only if *p*' and 'It is false that *p* if and only if not *p*' is quite sufficient: but it means facing the consequences of admitting that this is the *whole* explanation of the sense of these words, and this involves dethroning truth and falsity from their central place in philosophy and in particular in the theory of meaning.<sup>34</sup>

Dummett's argument brings out a tension between two ideas: the idea that the T-biconditionals explain the meaning of 'true' and the idea that meaning is to be explained in terms of truth conditions. If T-biconditionals are definitional of truth, if they explain what our understanding of 'true' consists in, then our understanding of 'true' presupposes a prior grasp of the meanings of the sentences of our language. Hence, truth cannot play a fundamental role in the theory of meaning; it cannot provide an explanation of our grasp of the meanings of sentences.

The tension is particularly vivid if one follows Donald Davidson and conceives of the theory of meaning for a language as a theory of truth for it.<sup>35</sup> The tension is now over two ways of reading the T-biconditionals: as elucidating the meanings of sentences and as elucidating 'true.' The two ways preclude each other. The former presupposes the concept of truth and uses the T-biconditionals to explain meaning; the latter presupposes meaning and uses the T-biconditionals to explain truth. By holding one element (truth or meaning) fixed, it appears, one can obtain the other. But one cannot use the T-biconditionals to extract both. As Horwich says, this is like having one equation and two unknowns.<sup>36</sup>

Fixing one unknown we can solve for the other, but we cannot solve for both simultaneously.

Notice that these considerations provide an argument against truth-conditional semantics only when they are supplemented with the full force of the Disquotation Thesis. A weaker thesis such as that the T-biconditionals are necessarily true is insufficient. Not only is there no tension between this weaker thesis and truth-conditional semantics, the very formulation of truth-conditional semantics requires a sense of 'true' for which the weaker thesis holds. Consider an arbitrary sentence '*p*' and an arbitrary possible situation *w*.<sup>37</sup> Truth-conditional semantics identifies the meaning of '*p*' with its truth conditions, say, *X*. Now suppose *w* is in *X*. The very formulation of truth-conditional semantics requires that there be a sense of 'true' on which '*p*' is true in *w*. Since *w* is in the truth conditions of '*p*,' the T-biconditional

'*p*' is true if and only if *p*

holds in *w*. By a parallel argument the biconditional holds also if *w* is not in *X*. Since *w* is arbitrary, the biconditional must be necessarily true. It follows that truth-conditional semantics requires a sense of 'true' on which the T-biconditionals are necessarily true.

The deflationary argument, if it is to work, requires the strong idea that the T-biconditionals explain the *meaning* or *sense* of 'true.' This suggests the following picture of our acquisition of 'true': We first learn some first-order words ('snow,' 'white,' etc.) and then we arrive at 'true' definitionally through the T-biconditionals.<sup>38</sup> Given this picture, it follows immediately that we cannot explain our understanding of 'snow is white' in terms of our understanding of 'true,' for our understanding of 'true,' according to the picture, rests on our prior understanding of 'snow is white.'

But now a basic difficulty with the argument comes into view. If anything like the above picture of the meaning of 'true' is correct, then an understanding of 'true' requires the possession of massive conceptual resources. For consider again the picture with which we are presented. We are told that we gain our understanding of 'true' through the T-biconditionals, that we acquire 'true' by laying down the totality of T-biconditionals as definitional of 'true.' But each biconditional plays



an important role in the resulting definition: It defines what it means to apply truth to one particular sentence. If some of the biconditionals are omitted, the result is at best a partial definition of 'true.' An individual who does not lay down some of the T-biconditionals as definitional of 'true' would have at best a partial notion of truth. To have a full notion of truth—to have a full understanding of the meaning of 'true'—requires, on this picture, a grasp of all the T-biconditionals. But this is possible only if the individual possesses all the concepts expressed by the terms in the right-hand sides of the biconditionals. Hence, on the above picture of the meaning of 'true,' a full understanding of 'true' is possible only for someone with massive conceptual resources.

An immediate response to this argument is to say that what defines 'true' is not the T-biconditionals, but something in their neighborhood—perhaps the form (T), or perhaps the general fact lying behind the T-biconditionals, or perhaps the rule of inference embodied in them. These suggestions are interesting but, as I shall argue in the next section, they do not provide a viable explanation of the meaning of 'true.' Furthermore, they cannot play the role that the Disquotation Thesis plays in Dummett's deflationary argument. Let us therefore for the moment set aside these suggestions and return to our original concerns: Should we think of the T-biconditionals as definitional of 'true'? Do the T-biconditionals explain what our understanding of 'true' consists in? In short, is the Disquotation Thesis true? Let us approach these questions indirectly. Let us ask: What are we denying in denying the Disquotation Thesis?

i. It is plain that we are not denying the T-biconditionals. Nor are we denying that the T-biconditionals are necessarily true. If the slogan "truth is a device of disquotation" is meant to say nothing more than this, then we are not denying the slogan. But the slogan so read does not provide a foundation strong enough to support deflationism.

ii. In denying the Disquotation Thesis, we need not even deny that there is a sense of 'definition' on which the T-biconditionals define 'true': We can accept the idea that the T-biconditionals fix the *extension*, and even the *intension*, of 'true'.<sup>39</sup> What we deny is that the T-biconditionals fix the *sense* of 'true'. When we evaluate a definition that aims to fix the extension or the intension of a predicate, we consider only how it carves the domain of its application into those objects that fall under the predicate and those that do not. The *ideology* of the definition, that is, the

totality of the concepts employed in the definiens of the definition, is entirely irrelevant.<sup>40</sup> So, the fact that the ideology of the T-biconditionals is vast does not cast any doubt on the idea that the biconditionals fix the extension and the intension of 'true.' But when we evaluate a definition that aims to capture the sense of a term, the ideology is of critical importance. For, the definition is now meant to capture what our understanding of the term consists in. If the definition is correct, a full understanding of the definiendum requires possession of the concepts in the definition's ideology. Let the *ideology* of a term consist of those concepts that are necessary and sufficient for an understanding of the term (assuming that there is such a totality).<sup>41</sup> Then, it is an adequacy condition on a definition that aims to capture the sense of a term that the ideology of the definition coincide with the ideology of the term. It follows that a definition that aims to capture sense may be inadequate simply because of the ideology that it employs.<sup>42</sup> This explains why the T-biconditionals are not an adequate definition of the sense of 'true.' If the T-biconditionals were adequate, then, given that their ideology is vast, it would follow that a full understanding of 'true' would require a massive repertoire of concepts. But, plainly, one can have a perfect understanding of 'true' even though one lacks, e.g., the concept of set or that of relativistic mass. The T-biconditionals fail to define the sense of 'true' because they attribute much too large an ideology to 'true.'

iii. In denying the Disquotation Thesis, we are not denying the observation that lies at the foundation of deflationism: that in asserting "'snow is white' is true" one typically asserts nothing more nor less than "snow is white." Deflationism goes on to explain this pragmatic fact in a certain way. And it is this explanation that we deny. According to deflationism, the pragmatic fact obtains because the sentences "'snow is white' is true' and 'snow is white' are synonymous, and the synonymy obtains because of the meaning of 'true.' Deflationism thus explains the pragmatic fact solely on the basis of the meaning of 'true.' But the deflationary explanation is not the only possible, or the most plausible, one. The pragmatic fact is sufficiently explained by the observation that in a typical situation the T-biconditional

(13) 'snow is white' is true if and only if snow is white,

is common knowledge, and indeed trivial common knowledge. Deflationism goes wrong because it reads the pragmatic fact into the very analysis of 'true.'

iv. This point is connected to the previous one. In denying the Disquotation Thesis we are not forced to deny that the T-biconditionals are

trivial. Nor are we forced to deny that there is a sense of ‘analytic’ on which the T-biconditionals are analytic. We can grant, for example, that a person who knows the meanings of all the parts of (13) will thereby know that

‘snow is white’ is true if and only if snow is white.

We can grant therefore that there is a sense in which the T-biconditionals are “true solely in virtue of meaning.”<sup>43</sup> But this is not to say that the T-biconditionals are “true solely in virtue of the meaning of ‘true’”; that someone who knew only the meaning of ‘true’ would thereby know the biconditionals. I think the point is important because the intuitive pull of the Disquotation Thesis comes from the seeming triviality and analyticity of the T-biconditionals. This makes us think that the T-biconditionals explain the meaning of ‘true’ and that an adequate definition of ‘true’ must imply the biconditionals (Tarski’s Convention T). But in thinking thus we make an unwarranted leap; a leap from common sense to deflationism.

In denying the Disquotation Thesis, then, we are not denying any of our commonsensical ideas about truth. We are denying a very specific claim about the meaning of ‘true,’ a claim that plays a key role in the deflationary argument from Dummett considered above. And we are denying a picture of how we arrive at our understanding of ‘true,’ a picture that makes the deflationary attitude compelling. Once we shed the claim and the picture nothing remains, I believe, to make plausible the deflationary attitude in the theory of meaning.

## V

The T-biconditionals make it tempting to believe that the concept of truth is simple, that a complete analysis of the meaning of ‘true’ is easily given. We readily grant that analysis of meaning is, in general, a difficult task. Even the meaning of such a simple word as ‘table,’ we believe, is difficult to specify. But when it comes to ‘true’ the T-biconditionals make it tempting to suppose that a reductive analysis of its meaning is possible. Even if we accept the point that an explanation of the meaning of ‘true’ should not employ the massive ideology of the T-biconditionals—and that therefore the T-biconditionals themselves do not explain the meaning of ‘true’—the thought persists that something in the neighborhood of

the T-biconditionals does explain it. What matters to the meaning of 'true,' we feel, is not the details of the particular T-biconditionals, but the general idea captured by them. We want to say that the meaning of 'true' is explained by the *form* (T),

(T) '\_\_\_\_' is true if and only if \_\_\_\_\_,

not by the particular biconditionals. And evidently the form does not carry with it a heavy ideology.

But how does a form explain the meaning of a predicate? This type of explanation of meaning is quite different from the usual sort. Form (T) does not explicitly state the application conditions of 'true' (or it would not have overcome the ideology problem), but this is what we expect from an explanation of the meaning of a predicate.<sup>44</sup> So how does (T) constitute an explanation of the meaning of 'true'? Several approaches suggest themselves as ways of answering this question. Let us examine a few. Our examination will cast doubt on the idea that a reductive analysis of 'true' is possible.

### (A) The Generalization Approach

This approach tries to make sense of the idea that (T) explains the meaning of 'true' by appealing to the general truth corresponding to (T). What explains the meaning of 'true,' on this approach, is not the totality of the T-biconditionals but the general fact that

(GT) All instances of the form (T) [i.e., all T-biconditionals] are true.

The suggestion has some attractive features. It explains the meaning of 'true' using a formula whose ideology is highly limited. Moreover, the formula is plainly analytic of the terms it involves. Anyone who understands the meaning of 'form (T),' 'true,' etc., must grant the truth of (GT).

Unfortunately, however, the suggestion faces an obvious but overwhelming problem: It explains the meaning of 'true' using a formula that itself involves 'true.' The circularity is not intrinsically objectionable.<sup>45</sup> But the particular form it takes here violates material aspects of the meaning of 'true': The proposal fails to yield the T-biconditionals. Imagine we give (GT) as an explanation of 'true' to someone who does not yet understand the word. This person will be able to deduce from (GT) that

(14) “‘Snow is white’ is true if and only if snow is white’ is true.

But how can he eliminate the last occurrence of ‘true’ and arrive at the T-biconditional

(15) ‘Snow is white’ is true if and only if snow is white?

To eliminate it he needs to derive the T-biconditional

(16) “‘Snow is white’ is true if and only if snow is white’ is true if and only if [‘snow is white’ is true if and only if snow is white].

But, again, (GT) does not yield (16) but only that (16) is true. A parallel difficulty blocks attempts to eliminate this new unwanted occurrence of ‘true.’ Our imaginary learner can derive of any T-biconditional that it is true, but he cannot derive the T-biconditional itself.

Note that if we *presuppose* the notion of truth, then the present strategy is a good way of spelling out the idea that a form explains the meaning of a predicate. The trouble is that the strategy works only if the meaning of ‘true’ is taken as given, not otherwise. The strategy cannot therefore be used to explain the meaning of ‘true.’ Rather, the meaning of ‘true’ is needed to make sense of the strategy.

## (B) The Syntactic Approach

This approach reads (T) as expressing a syntactic rule, a rule to the effect that a declarative sentence can be transformed by the addition (and deletion) of the marks,

(17) ‘        ’ is true,

without altering the sense of the original sentence.<sup>46</sup> The approach thus views addition of the marks (17) as analogous to the passive transformation. Both transformations have a limited utility, but are insubstantial modifications of the original.

The syntactic approach gets around the ideology problem. And, it, unlike the previous approach, avoids circularity in its explanation. However, it cannot be regarded as explaining the meaning of ‘true.’ At best, the approach explains the meaning of (17) *when (17) is viewed as one syntactic unit*; it does not explain the meanings of the parts of (17). The approach does not even entitle us to treat ‘is true’ as a predicate. It therefore does not explain the role ‘is true’ plays when it occurs in com-

bination with pronouns and general terms. If we follow the syntactic approach, we face problems explaining such simple inferences as the following:

The sentence Bill uttered is 'snow is white'.

'Snow is white' is true.

Hence, the sentence Bill uttered is true.

The approach does not entitle us to treat the occurrence of 'snow is white' in the second premiss as a singular term. Consequently, we cannot explain the inference as an instance of Leibniz's principle of indiscernibility of identicals.

### (C) The Inferential Approach

This approach uses the idea that the meaning of certain items in our language is specified by their inferential roles. The meaning of the truth predicate, it is suggested, is given by the rules of inference embodied in (T): to infer "'\_\_\_\_' is true' from '\_\_\_\_'; and, conversely, to infer '\_\_\_\_' from "'\_\_\_\_' is true.'

This approach to the explanation of meaning, while attractive for some parts of our language, is distinctly less attractive when applied to the truth predicate. For, if truth is explained in terms of inference, how do we explain our understanding of inference? How do we explain inference without appeal to the notion of truth? The natural response to the query is as follows. Inference is to be explained in terms of its role in our conceptual practices—practices of assertion, denial, supposition, verification, etc. These practices, the suggestion goes, are governed by various norms, and an explanation of inference will specify the role that it plays in these norms. Thus, our understanding of inference consists in understanding such things as that if  $q$  can be inferred from  $p$  then the assertion of  $p$  commits one to  $q$ , that one cannot assert  $p$  and also deny  $q$ , that a verification of  $p$  counts as a verification of  $q$ , etc.

Obviously, the suggestion is viable only if one can explain "assertion," "denial," "commitment," etc., without appeal to truth. But can this be done? How is one to make sense of our conceptual practices without any appeal to the notion of truth? And supposing sense can be made, do our practices make *more* sense when one allows the use of the notion of

truth? The inferential approach remains a large promissory note until it provides a satisfactory answer to these questions. I will not attempt to speculate on how the approach might be developed, but I would like to make two remarks about it.

First, the inferential approach to meaning does not need to forgo the notion of truth in order to stay true to its philosophical motive. *Use of the notion of truth in an explanation of our conceptual practices (and of meaning) doesn't immediately commit one to a referential picture of language.* It seems to me that the burden of explaining truth for a body of discourse in non-referential terms is lighter than that of giving a reductive explanation of the concept of truth. The inferential approach to meaning, it seems to me, takes a wrong turn when it denies itself the use of truth and takes on the burden of explaining the meaning of truth in inferential terms.

Second, even if truth does not play a substantive role in the explanation of our conceptual practices, it most likely does play an expressive role in their description. That is, truth is probably needed to describe basic facts about our conceptual practices—facts which are constitutive of them. For example, a description of our understanding of inference will, in all probability, need to mention our knowledge of the general fact of which the following is an instance.

If 'snow is white' can be inferred from 'everything is white', then snow is white if everything is white.

But how else can one express the general fact than by using the truth predicate: If a sentence *A* can be inferred from a sentence *B*, then *A* is true if *B* is? If this thought is correct, then the prospects for an inferential approach to the meaning of 'true' are bleak indeed.

None of the above approaches, then, are likely to yield a viable account of the meaning of 'true.' Let us observe also this: Even if these approaches were to lead to a viable account, it is an open question whether the resulting account would support any deflationary claims, and, if it did support some, it is an open question which ones it would support. For example, suppose that the inferential approach overcomes the obstacles in its way and offers an acceptable account of the meaning of 'true.' This

account will not, as far as I can tell, support a blanket deflationism in the theory of meaning. The account, plainly, could not play the role that the Disquotation Thesis played in the deflationary argument from Dummett considered in section IV. That argument rested on the idea that our understanding of “‘\_\_\_\_’ is true” presupposes an understanding of ‘\_\_\_\_’. The inferential approach, far from supporting this idea, is designed to overcome the problem that the idea creates. Further, the account the inferential approach proposes will, presumably, explain our understanding of ‘true’ in terms of our understanding of a limited range of terms, conceptual practices, etc. It is intuitively plausible that ‘true’ could not be used to provide an explanation of our understanding of the terms, practices, etc., within this range. But this allows truth to play an important role in an explanation of those terms and conceptual practices that lie outside the range. The account therefore will not make plausible a blanket deflationism in the theory of meaning.

Let us return to the original, disquotational, account of truth and take stock. The account, to review briefly, goes as follows: “The usefulness of truth lies in the expressive power that it provides. The truth predicate, by providing us with an indirect means of quantifying over sentence positions, enables us to express certain infinite conjunctions and disjunctions. To perform this function truth must be a device of disquotation: Applied to a quoted sentence it must undo the effect of the quotation marks. This function therefore requires the T-biconditionals to be definitional of ‘true’.” There are readings of this account on which it contains no errors, only insights. The key is how we understand “express,” “device of disquotation,” and “definition.” Suppose we understand them extensionally. Then the account reads (in part): “The generalizations involving truth are materially equivalent to the corresponding infinite conjunctions/disjunctions. To ensure this equivalence, truth needs to be a device of disquotation in the sense that the T-biconditionals need to be true. The T-biconditionals define ‘true’ in the sense that they fix the extension of ‘true’.” There is also an “intensional” reading of the disquotational account. This reading is parallel to the one just given, but it takes “express” to require necessary equivalence, “device of disquotation” to require necessary truth of the T-biconditionals, and “definition” to require the fixing of intension. The disquotational account, when it is



read in either of these ways, is good, true, and insightful. The only point that I have insisted on is that on these readings the account is not strong enough to carry the burden of deflationism.

Deflationists read (and, to sustain their conclusions, need to read) the disquotational account in very strong ways. Here is one such reading:<sup>47</sup> “The generalizations involving truth are abbreviations for (and, hence, mean the same as) the corresponding infinite conjunctions/disjunctions. To ensure this equivalence, truth needs to be a device of disquotation in the sense that guarantees the synonymy of “‘\_\_\_\_’ is true’ and ‘\_\_\_\_’. Thus, if ‘true’ is to perform its function, the T-biconditionals must be definitional in the sense that they explain the meaning of ‘true.’” The reading goes on to add: “This makes truth a simple concept. What it means, what our understanding of it consists in, and how we acquire it—these all have a simple explanation. The meaning of ‘true’ is given by the T-biconditionals, our understanding of it consists in our acceptance of the T-biconditionals, and we acquire it by laying down the biconditionals as its definition.”

The objection from ideology, given in section IV, puts in doubt each element of this reading. The T-biconditionals do not provide an adequate account of the meaning of ‘true’ because they impute to ‘true’ a massive ideology. The sentences “‘\_\_\_\_’ is true’ and ‘\_\_\_\_’ are not always synonymous, for the concepts needed to understand the latter are not necessarily needed to understand the former. The generalizations involving ‘true’ do not mean the same as the corresponding infinite conjunctions/disjunctions, for again the two do not involve the same conceptual resources.

It is remarkable that not only do the deflationary claims fail, but that an explanation of the usefulness of ‘true’ lies in their failure. An example: One important reason why generalizations involving ‘true’ are useful is precisely that they do *not* mean the same as the corresponding infinite conjunctions/disjunctions. Consider the generalization ‘all men are mortal’ for comparison. One reason why this generalization is useful is that it enables us to express a fact about all men without requiring of us the ability to say of each man that he is mortal. The generalization expresses, in a weak sense, the conjunction of its instances *without* being

synonymous with the conjunction. It thus enables us to express (weakly) a conjunction that we lack the resources to formulate explicitly—here lies its usefulness. The same holds for generalizations involving ‘true.’ One reason for their usefulness is that they are not synonymous with the corresponding infinite conjunctions/disjunctions. They allow us to express (weakly) these infinite conjunctions/disjunctions, even though our conceptual resources are meager.

More generally, one important source of the usefulness of ‘true’ is its remarkable double character: (i) that an understanding of ‘true’ requires only a limited range of conceptual resources, and yet (ii) ‘true’ enables us to talk and think about things that lie far beyond this range. (So, one reason ‘true’ is useful is precisely that the T-biconditionals do not define its sense.) This double character also constitutes, it seems to me, the most fundamental mystery of truth. The meaning of ‘true,’ like that of many other words, is difficult to explain; it becomes mysterious when we consider what ‘true’ enables us to do. ‘True’ appears simple to the deflationists, I think, because they overlook its most puzzling feature.

I have focused in this essay on homophonic truth (“true in English”) because it is here that the deflationist account appears most plausible. And I have tried to show that even here it fails. When we turn to heterophonic truth (“true in Inuit,” “true in such and such an infinitary language”) the problems facing deflationism become more vivid, but in essence remain the same. Suppose we have somehow gained ‘true in English,’ and suppose *L* is a language that can express things that are inexpressible in English (perhaps *L* is spoken by some alien creatures; perhaps *L* is an infinitary language that we find theoretically useful to talk about). How then can we gain ‘true in *L*’ when all we have to work with is ‘true in English’? We cannot say that a sentence of *L* is true iff it has a true translation in English, for this will make all untranslatable sentences of *L* untrue. How then will the explanation go?<sup>48</sup> Deflationism needs to explain ‘true in *L*’ without using the conceptual resources of *L*. The problem that must be solved is in essence the same as the fundamental problem we pointed to above. How to reconcile two features of ‘true in *L*’: (i) that it enables us to talk about the inexpressible contents of *L*, but (ii) the explanation of its meaning does not appeal to those contents.

In conclusion: Deflationists think that truth is a simple concept, one that has a simple analysis. The analysis the deflationists offer *is* simple, but, unfortunately, it makes truth far too complicated—it attributes to truth a vast ideology. We examined several attempts to get around this problem, but none resulted in a plausible account of the meaning of ‘true.’ Now we are left with questions: *What does our understanding of ‘true’ consist in? How can one explain the meaning of ‘true’ using a limited ideology?* It is a fact that we understand truth attributions even when truth is attributed to a sentence (or thought or representation) that lies beyond our conceptual resources. *What do we understand by such attributions?* We seem to grasp something general about what it is for a sentence (or thought or representation) to be true. *But what is it that we understand?* Once we overcome the spell of deflationism we are no longer inclined to brush these questions aside with simple answers. We regain our original sense that there is something very mysterious about truth and that an exploration of this mystery may illuminate the nature of our thought and our language.<sup>49</sup>

## Notes

1. See, for instance, Rudolf Carnap, *Introduction to Semantics* (Cambridge, Mass.: Harvard University Press, 1942).
2. Michael Williams, “Epistemological Realism and the Basis of Scepticism,” *Mind* 97 (1988): 415–39. The extract appears on 424.
3. I shall base my account of deflationism on the writings of a number of philosophers. I want to emphasize that while there are important similarities in the ideas of the philosophers I rely on, there are also important differences. No views, unless explicitly attributed to the individual authors, should be ascribed to them.
4. Richard Rorty, *Consequences of Pragmatism* (Minneapolis: University of Minnesota Press, 1982), xiii.
5. Scott Soames, “What Is a Theory of Truth?” *Journal of Philosophy* 81 (1984): 411–29. The extract appears on 429. (See chap. 17—Ed.)
6. Paul Horwich, *Truth* (Oxford: Basil Blackwell, 1990), 54.
7. Deflationists have offered several closely related descriptions of ‘true.’ In this paper I choose to focus on just one description—that contained in the disquotational account. Nevertheless, the arguments developed below apply in a straightforward way to many other deflationary descriptions. One notable exception is the strand of deflationism that relies on the Prosentential Theory of Truth of Dorothy Grover, Joseph Camp, and Nuel Belnap. A development of this strand

can be found in Dorothy Grover's essays in *A Prosentential Theory of Truth* (Princeton: Princeton University Press, 1992) and in Robert Brandom's "Pragmatism, Phenomenalism, and Truth Talk," in P. A. French, T. E. Uehling, Jr., and H. K. Wettstein, eds., *Realism and Antirealism: Midwest Studies in Philosophy 12* (Minneapolis: University of Minnesota Press, 1988). My view is that the Prosentential Theory has important insights into the logical grammar of truth. But these insights need to be supplemented with subsidiary theses before we can derive deflationary conclusions from them. I would want to argue that the subsidiary theses are problematic.

8. Although Quine's writings have provided much inspiration to the deflationists, a reasonable case can be made that Quine himself is no deflationist. First, the concept of truth seems to play a substantial role in Quine's philosophy of logic. Second, Quine takes a skeptical attitude towards many of the notions used in the defense of deflationism.

9. W. V. Quine, *Philosophy of Logic* (Englewood Cliffs: Prentice Hall, 1970), 12.

10. Stephen Leeds, "Theories of Reference and Truth," *Erkenntnis* 13 (1978): 111–29. The extract appears on 121.

11. The presence of ambiguity, context-sensitivity, self-reference, etc., in our language poses a challenge to the disquotational account. It forces us to recognize, for instance, that truth is not a simple predicate of sentences. I shall assume, for the sake of the argument, that the deflationists can meet the challenge. I shall often write as if the problematic elements are not present in our language. Also, when context allows it, I shall suppress relativity to language. I shall write 'true' in place of the longer 'true in English.'

12. I suppose I should call this thesis 'The Infinite Conjunction and Disjunction Thesis', but I want to save a few syllables.

13. See Horwich, *Truth*, 52 and 127. Recall also Quine's statement, "We need [a disquotational truth predicate] to restore the effect of objective reference when for the sake of some generalization we have resorted to semantic ascent" (emphasis added).

14. Sometimes the notion "T-biconditional" is understood in a wider sense. This allows a nonquotational name of a sentence to appear in the left hand side of the biconditional and a translation of the sentence to appear in the right hand side. Tarski constructed a definition (for certain languages) that implies the T-biconditionals in this wider sense. Since the definition implies the biconditionals, there could be no doubt that it was coextensive with truth. This refuted the skepticism of the Positivists. At the same time it made it seem that truth was a clear and simple notion. This paved the way for modern-day deflationism. John Etchemendy's paper, "Tarski on Truth and Logical Consequence," *Journal of Symbolic Logic* 53 (1988): 51–79, contains a good account of how Tarski's definition can be read in a deflationary way.

15. Observe that the mere truth (or even the necessary truth) of the T-biconditionals will not yield that disquotation is a “formal feature” of the truth predicate (Williams); nor will it yield the interdeducibility of the two sides of the T-biconditionals (Leeds); nor Quine’s claim that “[b]y calling the sentence [‘snow is white’] true, we call snow white.”

16. I put in the qualification “at least partially” because a full explanation of ‘true’ may require not only the T-biconditionals but also some such claim as ‘only sentences are true.’ I shall sometimes take the qualification as read and will not state it explicitly.

17. Horwich actually states this for a propositional notion of truth, but he wants to give a parallel account of the sentential notion. See *Truth*, 36–38, 52, 116, and 125.

18. Perhaps this explains why opponents of deflationism have been on the defensive in recent years.

19. Hilary Putnam accepted at one time an account of ‘true’ similar to the one sketched by Field. See his *Meaning and the Moral Sciences* (London: Routledge and Kegan Paul, 1978), 15–17. Putnam’s arguments against deflationism can be found in, among other places, his “On Truth,” in L. S. Cauman, I. Levi, and R. Schwartz, eds., *How Many Questions: Essay in Honor of Sidney Morgenbesser* (Indianapolis: Hackett Publishing Company, 1983), 35–56; and in “Does the Disquotational Theory Really Solve All Philosophical Problems,” *Metaphilosophy* 22 (1991): 1–13.

Christopher Hill is another nondeflationist who accepts parts of the disquotational account. See his ‘Rudiments of a Theory of Reference,’ *Notre Dame Journal of Formal Logic* 28 (1987): 200–219. I should note that Hill thinks that ‘refers’ and ‘true’ are ambiguous. He accepts the disquotational account only for one sense of ‘true,’ and he extends this account beyond the home language in a manner different from that of Field and Putnam.

20. Hartry Field, “Tarski’s Theory of Truth” (this volume, chap. 16). This reference is to the version in Mark Platts, ed., *Reference, Truth, and Reality: Essays on the Philosophy of Language* (London: Routledge and Kegan Paul, 1980), 104. The paper originally appeared in *Journal of Philosophy* 69 (1972): 347–75. This paper no longer represents Field’s present position. See his “The Deflationary Conception of Truth” (chap. 21—Ed.) in Graham MacDonald and Crispin Wright, eds., *Fact, Science, and Morality: Essays on A. J. Ayer’s Language, Truth and Logic* (Oxford: Basil Blackwell, 1986), 55–117. Field is now much more sympathetic to deflationism.

21. This causes a problem for any attempt to derive the strong reading of the Infinite Conjunction Thesis—and, more specifically, the synonymy of (2) and (4)—from the Disquotation Thesis. The Disquotation Thesis yields, we can grant, that (2) and (3) are synonymous. But to derive that (2) and (4) are synonymous we need the synonymy of (3) and (4), which unfortunately does not hold.

22. See his “Theories of Reference and Truth.”

23. See *Meaning and the Moral Sciences*.

24. Putnam rejects the idea that causal-explanatory laws are a reason to expect a physicalistic reduction. So, while Putnam thinks, *pace* Leeds, that truth is causal-explanatory, he rejects Field's quest for a physicalistic reduction of truth. See *Meaning and the Moral Sciences*, lectures 3–5.

25. Michael Williams, "Do We (Epistemologists) Need a Theory of Truth?" *Philosophical Topics* 14 (1986): 223–42. The extract appears on 232. The formulation of (6) that Williams is commenting on is this: "If we have true beliefs about how to attain our goals, we will generally attain them." Williams gives one other argument for not regarding (6) as a law. Since this argument does not rely on the disquotational account of 'true,' I shall not consider it here.

26. I am assuming here the Kripke-Putnam theory of reference.

27. Horwich, *Truth*, 45.

28. See 23–24 and 44–47 of Horwich's *Truth* for a fuller account.

29. Horwich, *Truth*, 52.

30. An analogy may make it clear that Horwich's argument is unsuccessful. Consider the generalization:

In aircrafts with autoland systems, accurate instrumentation promotes safe landings.

If Horwich's argument were successful then one could give a parallel deflationary explanation of this generalization and could argue that no further substantial explanation should be expected. But the deflationary explanation here is plainly unsatisfactory. An adequate explanation of why accurate instrumentation promotes safe landings would show how the actions of the autoland system are linked with the readings of the instruments and how these actions affect the flight behavior of the aircraft.

31. It seems to me that the Infinite Conjunction Thesis is false also if 'express' is taken to mean "interdeducible" or "warranted on the same occasions" (see the extract from Leeds' paper "Theories of Reference and Truth" given in section II). Let  $z$  be a set that contains sentences of a certain form. Then, the generalization 'all members of  $z$  are true' is not interdeducible with (nor is it warranted on the same occasions as) the infinite conjunction of the members of  $z$ . Neither the infinite conjunction nor the generalization carries information about what all the members of  $z$  are. But this is needed if we are to deduce one from the other.

32. To avoid misunderstanding let me say explicitly that I am not here defending a correspondence, or a physicalistic, or any other particular theory of truth. What I am defending is the claim that the meaning of 'true' does not make the search for a substantial theory of truth futile.

33. Michael Dummett, "Truth," in his *Truth and Other Enigmas* (London: Duckworth, 1978), p. 7. The essay originally appeared in *Proceedings of the Aristotelian Society* 59 (1959): 141–62. [Reprinted as chap. 10 of this volume—Ed.]

34. Dummett, *Truth and Other Enigmas*, 19. Dummett later rejected this conclusion. See the preface to *Truth and Other Enigmas*, especially xx–xxii.

35. For Davidson's views on the theory of meaning see his essays in *Inquiries into Truth and Interpretation* (New York: Oxford University Press, 1984). Davidson criticizes deflationism in "The Structure and Content of Truth," *Journal of Philosophy* 87 (1990): 279–328.

36. Horwich, *Truth*, 71.

37. Let us understand the variable  $p$  substitutionally in this argument.

38. Recall the extract from Field's paper "Tarski's Theory of Truth" given in section II.

39. I follow Carnap and think of "intension" as "extension-across-possible-situations." For a theory of how the T-biconditionals fix the extension (more precisely, the signification) of 'true,' even in the presence of vicious self-reference, see Nuel Belnap's and my *The Revision Theory of Truth* (Cambridge, Mass.: MIT Press, 1993).

40. Quine uses 'ideology' in an analogous, though not identical, way.

41. I shall understand "ideology of a concept" in a parallel way.

42. Here is a simple illustration. Consider the definiton,

$x$  is a human iff  $x$  is an animal with such and such a DNA structure.

Supposing that the 'such and such' is properly filled out, the definition fixes correctly the intension of 'human'—the ideology of the definition is irrelevant to this assessment. But when we consider whether the definition captures the sense of 'human,' the ideology is highly relevant. The fact that the ideology of the definition includes the concept "DNA structure" makes the definition implausible as an explanation of the sense of 'human.'

43. In this sense of 'analytic,' some analytic truths are open to substantial explanations. Contrast the analytic truths, "all bachelors are males" and "'snow' means snow." It is unreasonable to expect a substantial answer to the question "Why are all bachelors males?" but not to the question "Why does 'snow' mean snow?"

44. I am setting aside circular predicates here.

45. As Belnap and I have argued in *The Revision Theory of Truth*.

46. Mark Wilson and Eric Dalton, independently, suggested this approach to me. They do not endorse it.

47. This is not the only possible strong reading, but it will highlight the points made earlier.

48. This kind of problem rules out the most obvious deflationist response to the ideology objection. According to the response, each of us learns "true" first, not as it applies to English, but as it applies to one's own personal idiolect. That is, one first acquires the concept "true-in-my-present-idiolect" and then using it acquires the full-fledged "true." The problem of explaining how one goes from

“true-in-my-present-idiolect” to “true” seems to me to be much harder than that of explaining “true” using a limited ideology. The response reduces a very hard problem to a virtually impossible one.

49. I presented some of the ideas of this paper in talks at the University of Minnesota, Boston Colloquium for the Philosophy of Science, University of Delhi, and Indiana University. I wish to thank my auditors (especially André Chapuis, John Etchemendy, Geoffrey Hellman, Paul Horwich, Ranjit Nair, and Scott Soames) for their helpful comments, suggestions, and queries. Members of my Spring 1993 Metaphysics Seminar at Indiana University also helped me—not only through their ideas and suggestions, but also through their friendly skepticism towards my favorite arguments. I do not want to reproduce the whole roster for the seminar, but I should mention David Chalmers, Eric Dalton, Craig DeLancey, Jim Hardy, Ingo Farin, Adam Kovach, Malcolm MacIver, Gregg Rosenberg, and Jerry Seligman. I owe a special debt to Nuel Belnap, Dorothy Grover, Jerry Kapus, and Mark Wilson. Over the years I have had, and have benefited from, numerous discussions with these philosophers. None of them, of course, should be held responsible for the flaws of this essay. I know Dorothy will disagree with much of what I say. My views are, I think, closest to those of Jerry Kapus. See his “Truth and Explanation” (Ph.D. diss., University of Illinois at Chicago, 1992). Finally, I wish to thank Marian David, Allen Hazen, and Chris Hill for their comments on this essay.





## A Defense of Minimalism

---

Paul Horwich

My aim in this paper is to clarify and defend a certain “minimalist” thesis about truth: roughly, that the meaning of the truth predicate is fixed by the schema, “The proposition that  $p$  is true if and only if  $p$ .”<sup>1</sup> The several criticisms of this idea to which I wish to respond are to be found in the recent work of Davidson, Field, Gupta, Richard, and Soames, and in a classic paper of Dummett’s. But before addressing these criticisms, let me begin by saying something more about the thesis itself.

Consider biconditionals like

$\langle \text{snow is white} \rangle$  is true  $\leftrightarrow$  snow is white

and

$\langle \text{lying is wrong} \rangle$  is true  $\leftrightarrow$  lying is wrong<sup>2</sup>

—that is, instances of the *equivalence schema*:

$\langle p \rangle$  is true  $\leftrightarrow p$

It can be argued that such biconditionals are *epistemologically fundamental*: we do not arrive at them, or seek to justify our acceptance of them, on the basis of anything more obvious or more immediately known. In addition, it can be argued that our underived inclination to accept these biconditionals is the source of *everything else* we do with the truth predicate. For example, from the premises

What he said is that he was abducted

and

What he said is true

we are prepared to infer

He was abducted

This particular use of the word “true” is explained by supposing that we first employ Leibniz’s Law to get from our pair of premises to

⟨He was abducted⟩ is true

and then invoke the relevant instance of the equivalence schema. And, more generally, it can be made plausible that no further fact about the truth predicate—nothing beyond our allegiance to the equivalence schema—is needed to explain *any* of our ways of using it. It is for this reason that we are entitled to conclude that the meaning of “true” is determined by that schema. For, plausibly, the property of a word that constitutes its having the particular meaning that it has should be identified with the property that explains the *symptoms* of its possessing that meaning—and these symptoms are the various characteristic ways in which it is used.<sup>3</sup> Thus my minimalist thesis is the product of two prior claims: first, that our underived endorsement of the equivalence schema is explanatorily fundamental with respect to the overall use of the truth predicate, and second, that the meaning of any word is engendered by the fact about it that explains its overall use.<sup>4</sup>

This line of thought can be challenged at various points and no doubt stands in need of considerable further justification.<sup>5</sup> But my main aim here is not to defend my *route* to the minimalist conclusion but rather to defend that conclusion itself: namely, that the meaning of “true” stems from the equivalence schema. For most of the recent objections to this thesis do not target any particular rationale for it, but purport to demonstrate that the thesis itself cannot be correct. However, before addressing these objections, let me help to prepare the ground for my replies to them by saying a little more to clarify just what the proposal is, and is not, intended to encompass.

Several different kinds of theory, with very different explanatory objectives, might appropriately be labeled “theories of truth.” So it is important to be clear about what sort of theoretical work the minimalist proposal is not meant to do and should not be blamed for failing to do. In the first place, it is not intended to provide an *explicit definition* of the word “true,” either descriptive or stipulative. Therefore, it does not offer a way of rearticulating the contents of sentences containing the word. Indeed, it implies that no such reformulations are possible.<sup>6</sup> In the

second place, the proposal does not amount to a substantive *reductive theory* of the property of being true—something in the style of “Water is H<sub>2</sub>O”—which would tell us how truth is constituted at some underlying level. Again, it suggests that the search for such a theory would be misguided.<sup>7</sup> And in the third place, the proposal is not a “theory of truth” in the sense of a set of fundamental theoretical postulates on the basis of which all other facts about truth can be explained. Its immediate concern is with the word “true” rather than with truth itself. It purports to specify which particular nonsemantic fact about that word is responsible for its meaning what it does; and the fact it so specifies is our underived allegiance to the equivalence schema.

Now let me turn to an array of objections. I will look at two difficulties raised by Donald Davidson, one posed by Hartry Field, three devised by Anil Gupta, one due to Mark Richard, one that I put to myself, and one old, but still influential, objection of Michael Dummett’s. My discussion of each of these problems will be fairly brief—merely indicating the lines along which I think the response should be given, rather than giving it in full.

*Objection 1* The minimalist proposal implies that the meaning of “true” is established on the basis of the meaning of “the proposition that...” For someone’s acceptance of, for example, “The proposition that *Hesperus rotates* is true if and only if *Hesperus rotates*” manifests a standard understanding of the truth predicate only to the extent that the component expression “the proposition that” is being understood in the standard way. Thus minimalism implies that one must already understand “that” clauses—i.e., one must understand sentences of the form “*u* expresses (i.e., means) the proposition that *p*”—in order to be in a position to acquire the concept of truth. But this surely gets things the wrong way round! Surely the intimately related notions of meaning and proposition must be analysed *in terms of* truth. Otherwise, we wouldn’t be able to account for the *compositionality* of meaning. More specifically, we must (for the sake of explaining compositionality) suppose that

*u* expresses the proposition (i.e. means) that *p*  
consists in the fact that

*u* is true (i.e., expresses a truth) if and only if *p*

Thus truth is conceptually prior to meaning, contrary to what is required by the minimalist proposal. (Davidson<sup>8</sup>)

The compositionality of meaning is the fact that the meanings of sentences depend on the meanings of their component words and on how those words are put together. Or, to put it another way, it is the fact that our interpretations of the complex expressions of someone's language derive from our interpretations of that person's primitive terms. Now, despite the considerations advanced by Davidson in his influential essay "Truth and Meaning" and elsewhere, this fact does not really call for an analysis of meaning in terms of truth conditions.<sup>9</sup> A viable alternative to the Davidsonian strategy is to endorse a Fregean formulation of the principle of compositionality and to regard this as not susceptible to explanation. Thus we would take it as explanatorily fundamental that whenever a complex expression is formed by applying a function expression (e.g., a predicate) to a sequence of argument expressions (e.g., names) the meaning of the complex is the result of applying the meaning of the function expression to the meanings of its arguments. For example, it would be explanatorily fundamental that the meaning of the result of applying the predicate "rotates" to the name "Hesperus"—i.e. the meaning of that sentence—is the meaning of "rotates" applied to the meaning of "Hesperus." On the basis of this unexplained Fregean principle and from specifications of the meanings of the words in a language, it is possible to deduce the meanings of every sentence (structurally described) and hence to interpret the entire language.<sup>10</sup>

Davidson's objection to the Fregean view of compositionality is that it is "vacuous": that it implies merely that "Hesperus rotates" means whatever it means, that it fails to show how that meaning depends on the structure of the sentence, and that it does not help us to give interpretations of complex expressions. But these criticisms are wrong. The Fregean principle can be no emptier than the compositionality it expresses; it says that "Hesperus rotates" means the result of applying the meaning of "rotates" to the meaning of "Hesperus," which is not a tautology; it does offer a structure-dependent characterization of what the sentence means; and for the sake of interpretation, such characterizations are all that are needed. Granted, the Fregean proposal is more of an *articulation* of compositionality than an *explanation* of it. But there is no reason to suppose that compositionality can or should be explained. No deeper account is needed for purposes of interpretation, i.e., for enabling de-

ductions of the meanings of complexes on the basis of the meanings of their parts.

Moreover, this account of compositionality puts no constraint at all on how the meanings of *words* are in turn constituted. In particular, it squares perfectly well with supposing that the meanings of words are engendered by non-truth-theoretic aspects of their use. Indeed, I believe such a view of word meaning can be made highly plausible.<sup>11</sup> In that case, our understanding of (and ability to deploy) “*u* means *that Hesperus rotates*” will emerge from our knowledge of the uses (and hence meanings) of “Hesperus” and “rotates” and from our appreciation of how the sentence results from applying one of them to the other. Thus, contrary to this initial objection of Davidson’s, it is quite reasonable to allow that we could first possess the concepts of meaning and proposition and then, on that basis, fix the meaning of the word “true” (and acquire the concept of truth) by accepting instances of the equivalence schema.<sup>12</sup>

*Objection 2* “That” clauses cannot be regarded as referring terms, because there is no way of seeing how their referents would be determined by the referents of their component words. Therefore, sentences like “The proposition *that Hesperus rotates* is true,” insofar as they are construed as predicating truth of the propositions to which the “that” clauses refer, are in fact unintelligible. But if such truth ascriptions (so construed) are unintelligible, then the minimalist proposal cannot be correct. (Davidson<sup>13</sup>)

Davidson’s basic reason for maintaining that alleged referents of “that” clauses would not be determinable by the referents of the parts of these clauses is—following Frege—that substitution of coreferential terms (e.g. putting “Phosphorus” in place of “Hesperus”) within a “that” clause occurring in some sentence (e.g., “John believes that Hesperus rotates”) will not always preserve the truth value of that sentence. But why does he not continue along Fregean lines and conclude that an expression within a “that” clause does not have its *standard* referent but instead refers to the *meaning* of that expression? Why not identify the referent of “the proposition that Hesperus rotates” with the meaning of “Hesperus rotates” and identify the referents of the contained words “Hesperus” and “rotates” with *their* meanings? And why not suppose, as suggested in my response to objection 1, that the meaning of “Hesperus rotates” is the result of

applying (in the sense of applying a *function* to an *argument*) the predicative meaning of “rotates” to the nominal meaning of “Hesperus”?

Davidson has made two objections to this Fregean proposal. His first objection, given in “Truth and Meaning” and discussed above, is that it is unilluminating. But, as we have seen, this criticism is overstated. No more substantive account of compositionality than is contained in the Fregean principle is required to derive interpretations of complex expressions from interpretations of their parts. Admittedly, on this account, compositionality is left unexplained—it is treated as an explanatorily basic phenomenon. But that may well be the correct view of it: there may well be much less to the so-called “problem” of compositionality than is often supposed.

More recently, in his essay “The Folly of Trying to Define Truth,” Davidson has advanced a somewhat different reason for rejecting the Fregean picture (and hence for concluding that the referents of the parts of a “that” clause could not determine a referent for the whole). He observes that the meanings of words in “that” clauses are just their normal meanings. After all, we understand “The proposition that Hesperus rotates” only if we understand the isolated sentence “Hesperus rotates.” And in the biconditional “The proposition *that Hesperus rotates* is true  $\leftrightarrow$  Hesperus rotates,” the two occurrences of “Hesperus rotates” are clearly supposed to be understood in the same way. But in that case, since meaning determines reference, how could words in “that” clauses fail to have their standard referents? And if they do have their standard referents, then “that” clauses cannot refer, since what would be determined by those standard referents would be the *wrong* thing (e.g., “that Hesperus rotates” would acquire the same referent as “that Phosphorus rotates”).

But two reasonable responses are available to this line of thought. First, we might well deny that meaning determines reference. We might suppose, on the contrary, that the referent of a term is fixed in part by the context in which it occurs. This general idea is quite uncontroversial. More specifically, we might say that the single meaning of “Hesperus” yields one referent (the planet) for standard (non-opaque) occurrences of the word, and that it yields a different referent (the meaning, or sense, of “Hesperus”) for occurrences of the word within “that” clauses.

An alternative response would be to deny that the referent of a complex expression is determined by the referents of its *grammatical* parts. We might prefer to say instead that the referents only of *logically articulated* expressions are determined by the referents of their parts. For we might suppose that although the words “Hesperus” and “rotates” are indeed used in the superficial grammatical form of “the proposition that Hesperus rotates,” the underlying logical form might be something in which those words are not, properly speaking, *used* but are merely mentioned (e.g., “the proposition expressed by the sentence ‘Hesperus rotates,’ as I currently understand it”). Admittedly, this diverges from Frege. Nonetheless, it would treat “that” clauses as singular terms, and it would conform to Davidson’s requirements on such a treatment, namely, that their referents be determined by the referents of their logical parts, and that these parts have the same meanings inside “that” clauses as they do outside.<sup>14</sup>

*Objection 3* The minimalist proposal would leave it mysterious how we are able to attribute content to sentences that predicate truth of foreign statements we can’t understand. For if an utterance *u* is known to mean (say) *that dogs bark*, then (according to the proposal) the sentence “*u* is true” (or “*u* expresses a truth”) might be interpreted as saying roughly *that dogs bark*, whereas if *u* is not understood, then the proposal enables us to attach no meaning at all to “*u* is true.” But we surely do think that it is meaningful to predicate truth of statements we cannot understand. So the minimalist proposal is defective. (Field<sup>15</sup>)

But, as we have already seen, the proposal is not intended as an explicit or contextual definition of “true”: it does not purport to provide a way of reformulating or rearticulating the content of each sentence containing the word “true.” On the contrary, it implies that such a thing is not possible. So one cannot reasonably complain that the minimalist proposal fails to yield a conceptual analysis of the sentence “*u* is true.” One certainly *could* complain if the proposal implied that this sentence lacked content. But it has no such implication. It aims to specify the underlying use property in virtue of which the truth predicate means what it does. To that end, it identifies certain tokens of that predicate as explanatorily fundamental and hence as meaning-constituting—namely, those that appear in instances of the equivalence schema. But other tokens of that



type may perfectly well have the very same meaning, as long as their deployment stems from the fundamental ones. Thus someone who reasons inductively to “*u* expresses something true” on the basis of the fact that the other assertions of the speaker—those that *can* be translated—have turned out to be true, uses the truth predicate with a constant meaning, one that is engendered by the equivalence schema.

*Objection 4* The equivalence schema is not enough to fix the meaning of the truth predicate, because exactly parallel schemata are satisfied by predicates that do not mean the same as “true.” For example, instances of the schema, “ $\langle p \rangle$  is true and not red  $\leftrightarrow p$ ” are just as obviously correct as instances of the equivalence schema. But “is true and not red” is not a synonym of the truth predicate. More generally, the schema “ $\langle p \rangle$  is  $f \leftrightarrow p$ ” will be endorsed relative to a variety of predicates “ $f$ ” with different meanings from one another. Therefore, for any given “ $f$ ,” it cannot be that our acceptance of instances of “ $\langle p \rangle$  is  $f \leftrightarrow p$ ” is what fixes the meaning of “ $f$ .” (Gupta<sup>16</sup>)

Indeed. However, according to minimalism, what fixes the meaning of the truth predicate is *not* merely our *allegiance* to the equivalence schema but, in addition, the fact that this allegiance is the use that is epistemologically basic—i.e., the fact that our endorsement of the equivalence schema is that use of “true” which is not derived from any more fundamental assumptions formulated by means of the truth predicate. Certainly there are parallel schemata, constructed with other predicates in place of “true,” that are equally acceptable. However, in every such case our commitment to the parallel schema is *derived*. For example, it is as a consequence of our endorsing the equivalence schema for truth, and of our accepting “No proposition is red,” that we accept instances of “ $\langle p \rangle$  is true and not red  $\leftrightarrow p$ .” Only in the case of the *truth* predicate does the corresponding schema capture what is epistemologically basic in our usage of the predicate. And it is this property that engenders its meaning.

*Objection 5* The minimalist proposal implies *either* that the word “true” will never be fully understood *or* that the meaning of each person’s truth predicate depends on, and varies with, whatever else is in his vocabulary. For the proposal is tantamount to the following definition:

$x$  is true  $\equiv [x = \langle \text{dogs bark} \rangle \ \& \ \text{dogs bark}; \text{ or}$   
 $x = \langle \text{pigs fly} \rangle \ \& \ \text{pigs fly}; \text{ or}$   
 $\dots \text{ and so on}]$

Therefore, if the “and so on” is intended to cover *all* propositions, then since some of them must involve concepts that no one possesses, the meaning of “true” will not be fully known to anyone. And if, alternatively, the definition of each person’s truth predicate is supposed to cover only those propositions he can currently grasp, then as new concepts are deployed and new terms coined, his definition of “true” will change. But neither of these alternative implications of minimalism is acceptable. Surely our understanding of the truth predicate is both complete and constant. (Gupta<sup>17</sup>)

Agreed. But my proposal is perfectly consonant with such intuitions, because it is not at all equivalent to the above definition. As already emphasized, the minimalist thesis does not offer *any* explicit definition. Rather, it purports to specify the fact of usage that provides the truth predicate with its meaning. That fact of usage, it claims, is our underived inclination to accept instances of the equivalence schema—a fact that remains the same while the rest of our language evolves. So, for example, at the moment that the term “tachyon” enters our language, we become inclined to accept

〈tachyons go backwards in time〉 is true  $\leftrightarrow$  tachyons go backwards  
in time

But this is merely one more application of a single and invariable regularity—our inclination to accept any instance of the schema that we understand. That inclination preceded the introduction of the term “tachyon” and was in no way altered by it. Thus the minimalist thesis does not imply that the meaning of the word “true” can’t be fully grasped, or that it changes with the expansion of our vocabulary.

*Objection 6* Our reliance on the equivalence schema will not suffice to explain our knowledge of *general* facts about truth. Consider, for example, “All propositions of the form  $\langle p \rightarrow p \rangle$  are true.” No doubt our particular logical convictions, together with our commitment to the equivalence schema, can explain, *for any single proposition*, why we take it to be true that this proposition implies itself. Thus we can explain, given our logical commitment to “dogs bark  $\rightarrow$  dogs bark,” why we also accept “The proposition *that dogs bark  $\rightarrow$  dogs bark* is true.” But we have not thereby explained how the above *generalization* is reached. Thus our allegiance to the equivalence schema does not really suffice to account for *all* uses of the truth predicate. Therefore, that practice does

not fix the meaning of “true,” contrary to what the minimalist maintains. (Gupta, Soames<sup>18</sup>)

Granted, some further explanatory premise is needed if every use of “true” is to be accounted for. But this concession provides an objection to the proposal only if the needed additional premise specifies properties of the word “true.” For only then will it emerge that our commitment to the equivalence schema, together with facts that have nothing specifically to do with the truth predicate, are insufficient to explain its overall use. But actually it is far from obvious that the premise we should add will explicitly concern the truth predicate.

Suppose, for example, it were the case that whenever someone can establish, for each  $F$ , that it is  $G$ , he then comes, on that basis, to believe that every  $F$  is  $G$ . Combined with such a fact (which does not explicitly concern the truth predicate) our disposition to accept, for each proposition of a certain form, that it is true would suffice to explain our acceptance of the generalization “Every proposition of that form is true.”

Of course, this response to the objection will not do as it stands, because the proposed extra explanatory premise is clearly incorrect. It is *not* always the case that the ability to establish, regarding each  $F$ , that it is  $G$ , engenders the belief that all  $F$ s are  $G$ . For example, suppose that someone mistakenly suspects that there are planets within the orbit of Mercury. In that case he might nevertheless be able to show, of every planet, that its distance from the sun is not less than Mercury’s, but he does not believe the generalization that all the planets have this property.

Obviously, he is not tempted to draw the generalization, because, although in fact he can establish for each  $F$  that it is  $G$ , he does not appreciate that he can—indeed, he would deny it. This suggests that a more plausible version of our extra premise would run along the following lines:

Whenever someone can establish, for any  $F$ , that it is  $G$ , *and recognizes that he can do this*, then he will conclude that every  $F$  is  $G$ .

It seems to me that this is more or less what we need to explain our acceptance of the generalization about truth. We can establish, for any proposition of the form  $\langle p \rightarrow p \rangle$  that it is true. Moreover, we are aware of this ability. For we see that any of the propositions in question may be

designated by some expression of a certain form. And we see that there is a general strategy which, given any such expression, will establish the sentence that results from writing the expression before “is true.”<sup>19</sup>

Thus we have a plausible explanatory premise that, in conjunction with our endorsement of the equivalence schema, will enable us to explain the acceptance of generalizations about truth. And since that premise does not explicitly concern the truth predicate, the need for it does nothing to suggest that the basic regularity governing the truth predicate has to go beyond our underived commitment to the equivalence schema.<sup>20</sup>

*Objection 7* Certain people (mostly philosophers) do not have a completely general inclination to endorse the equivalence schema. They hold, for example, that ethical pronouncements, contingent statements about the future, applications of vague predicates to borderline cases, or sentences containing empty names fail to yield acceptable instances. But these people nevertheless mean the same thing as we do by the word “true.” After all, we might disagree with them on the question “Can ethical pronouncements can be true?”, yet each of us expresses the issue in just that way. Consequently, it cannot be that to understand the English truth predicate, one must have an inclination to accept *all* instances of the schema. (Richard<sup>21</sup>)

Notice that the present objection is not the claim that certain instances of the equivalence schema are incorrect or the claim that some of them are unhesitatingly rejected by everyone.<sup>22</sup> The point rather is that *if* someone has no inclination to accept a certain class of instances, he might nonetheless understand them exactly as we do. And so we have to conclude, it would seem, that a *general* endorsement of the schema is *not* what provides the word “true” with its meaning.

But this conclusion is unwarranted. For we can invoke *social externalism* to accommodate our minimalist thesis to the facts under consideration. We might suppose that the tendency of some people to restrict the equivalence schema stems from confusion about the meaning of the truth predicate. We might suppose that their basic use of it does not quite match its meaning-constituting use, but that they are nonetheless correctly ascribed the standard concept of truth because they belong to a linguistic community in which that concept is the “right” one to have—the one deployed by the appropriate “experts.” Moreover, a good case

can be made that it is indeed those who restrict the schema, rather than those of us who do not, who are confused and mistaken. For they tend to be in the grip of the idea that truth is a substantive property, analysable in terms of “correspondence with facts.” They do not appreciate the real role of the truth predicate as a device of generalization—a role that requires that it be applicable to *any* proposition (i.e., to what is expressed by *any* declarative sentence). Thus one can say that the meaning-constituting fact about the truth predicate is the fact that explains the overall use of it by those who are not under the spell of a misbegotten philosophical theory. And this fact is our allegiance to the fully general equivalence schema.

*Objection 8* But couldn't we come across someone who denies that there is such a thing as truth, someone who is not prepared to accept *any* instances of the equivalence schema? Couldn't such a person nonetheless *understand* our talk of truth and nonetheless mean what we do by the word “true” when he says, for example, “Even though dogs bark, it is not true that they do”?

I believe this is correct. Moreover, social externalism with regard to meaning does not suffice to reconcile minimalism with such a scenario. For our imagined sceptic's use of the word “true” is not even *approximately* a matter of endorsing the equivalence schema. The moral would seem to be, rather, that there must be some use of “true” that (a) is implicit in, but weaker than, an endorsement of the equivalence schema, (b) is displayed by the sceptics and by ourselves, and (c) constitutes what we both mean by that word.

This conclusion is also suggested by the plausibility of allowing, contrary to what we have been supposing so far, that our endorsement of the equivalence schema is *not* epistemologically fundamental. For it would seem, rather, to be the product of two deeper factors: first, there being *some* term “*f*” such that all instances of an equivalence schema involving that term are accepted as basic (i.e., all instances of “ $\langle p \rangle$  is  $f \leftrightarrow p$ ”), and second, the conditional decision that *if* there is such a term, it shall be the word “true.” Given this way of factorizing our endorsement of the equivalence schema, it is plausible that although that endorsement is *sufficient* to fix the meaning of “true,” it is not *necessary*. Rather, all that is necessary is the second of these factors—the conditional commitment

to express the concept that satisfies the schema, *if any concept does*, using the word “true.” In that case, someone might reject the antecedent of the conditional—he might deny that there is any concept satisfying the schema—and yet agree with us about what the truth predicate means.<sup>23</sup>

Thus our initial minimalist proposal must be revised. Meaning what we do by the truth predicate is constituted *not* by an inclination to accept instances of the equivalence schema but rather by the commitment to have that inclination, on condition that one is inclined, for some “*f*” to endorse “ $\langle p \rangle$  is  $f \leftrightarrow p$ ”.

*Objection 9* Truth is valuable: we ought to pursue it, and we ought to avoid false belief. But these sentiments are not contained in (nor can they be extracted from) instances of “ $\langle p \rangle$  is true  $\leftrightarrow p$ ,” which are entirely non-normative. Consequently, our concept of truth is not fully captured by the equivalence schema. So the minimalist proposal is false. (Dummett<sup>24</sup>)

On the contrary, the equivalence schema *is* able to account for the normative force of truth. To see this, consider the specific norm of belief:

One should believe that wombats fly  $\leftrightarrow$  wombats fly

Clearly our commitments to norms like this one have nothing to do with the concept of truth, for that concept is completely absent from their articulation. Not, of course, that there is no call to explain *why* we make such commitments. The point is merely that one should not expect the concept of truth to be doing any of the explanatory work.

Let us, then, imagine that all such specific norms of belief are somehow explained.<sup>25</sup> Suppose, that is, we can account for our attachment to all norms of the form:

One should believe *that*  $p \leftrightarrow p$

Given the equivalence schema we will then be able to explain our attachment to every norm of the form

One should believe  $\langle p \rangle \leftrightarrow \langle p \rangle$  is true

that is, to every norm of the form

One should believe  $x \leftrightarrow x$  is true.

But this engenders (via the mechanism discussed in my response to objection 6) a commitment to the generalization

$(x)(\text{one should believe } x \leftrightarrow x \text{ is true})$

—or, in English, to the principle

One should believe what is true and only what is true.<sup>26</sup>

Thus the value we attach to true belief is explained by the role of truth as a device of generalization, which is itself perfectly explained by the equivalence schema.

The *full* minimalist picture of truth includes considerably more than the thesis I have been defending in this essay. It involves, besides the present claim about how the meaning of “true” is constituted, an affiliated view about the *function* of the truth predicate (namely, as just illustrated, that it is merely a device of generalization), an affiliated view about the underlying nature of truth (namely, that there is no such thing), and an affiliated view about the general shape of the basic theory that will best explain all the facts about truth (namely, that its postulates are instances of the equivalence schema). I have not attempted to elaborate or establish these further minimalist doctrines. However, since what I have been concerned with here is the central component of minimalism, my defense of that thesis, if successful, provides important support for the view as a whole.

## Notes

I would like to thank Hartry Field and Michael Lynch for their comments on a draft of this paper.

1. A note on the relationship between *deflationism* and *minimalism* about truth. Deflationism is the somewhat vague idea that truth is not a “substantive” property, that no reductive theory of it should be anticipated, and that our grasp of the truth predicate comes from our appreciation of how each statement specifies its own condition for being true. But philosophers who sympathize with this general point of view disagree among themselves about how best to elaborate it. Minimalism is one such strategy—the one defended here and previously articulated in my book *Truth* (2nd edition, Oxford University Press, 1998). Besides minimalism, the main alternative forms of deflationism about truth are (1) *disquotationalism*, according to which sentences (rather than propositions) should be regarded as the bearers of truth, and the schema “ $p$ ” is true  $\leftrightarrow p$  will be what defines the truth predicate; (2) *prosententialism*, which stresses the use of “That is true” to save having to repeat what has just been asserted and which denies that “true” should be logically formalized as a predicate; (3) *the redundancy theory*, whereby “The proposition that  $p$ ” means exactly the same as “ $p$ ”; (4) *the quan-*

*tificational theory*, which analyses truth talk in terms of substitutional quantification into sentence positions: “ $x$  is true” means “ $(\exists p)(x = \langle p \rangle \ \& \ p)$ ”; and (5) *Tarski’s theory*, which explains the truth of each sentence of a language in terms of the referential properties of its components and the logical structure in which they are embedded. The relative advantages of the minimalist version of deflationism cannot be fully spelled out here. But in a nutshell, its merits are (a) that it deals with our *actual* concept of truth, rather than some allegedly superior one; (b) that it does not attempt to explain truth in terms of notions that should themselves be explained in terms of truth (e.g., substitutional quantification); (c) that it recognizes that there is no call for an explicit definition; (d) that it can countenance the attribution of truth to propositions whose logical forms we do not know; (e) that by acknowledging truth as a property, it squares perfectly with its role as a device of generalization.

2. Here “ $\langle p \rangle$ ” abbreviates “the proposition that  $p$ ,” and “ $\leftrightarrow$ ” is the *material* biconditional.

3. This view of how meaning-constituting properties are to be identified is an instance of the general idea that an underlying property  $U$  constitutes a relatively superficial property  $S$  when  $U$ ’s being coextensive with  $S$  explains why possession of  $S$  has the symptoms that it does. My speaking of  $S$  as being *constituted* by  $U$  in those circumstances, rather than as being *identical* to  $U$ , rests on a fine-grained conception of “property” whereby two predicates stand for the same property only when they have the same meaning. This way of speaking does not preclude also deploying a more coarse-grained conception whereby constitution would suffice for identity.

4. I am inclined to say that a word’s meaning-property (e.g., “ $w$  means TRUE”) is *constituted by* (or *reduces to*, or *may be analysed as*) a certain way of using the word. But in light of the *relational* character of meaning properties, this thesis is in tension with the principle that any analysis of a complex property must result from the analysis of (at least one of) its constituents. One strategy here is to criticize that principle. Thus the Frege-Russell reduction of “the number of dogs owned by  $x = 0$ ” to “ $\neg(\exists y)(y \text{ is a dog owned by } x)$ ” is a plausible counterexample. Another strategy is to retreat to a weaker thesis. Rather than speaking of constitution (or reduction or analysis), we might say that the some usage of a word *engenders* (or *determines* or *is the explanatory basis for*) its meaning property.

5. For justification of the claim that the equivalence schema is explanatorily fundamental, see my book *Truth* (1998). For justification of the use theory of meaning, see my book *Meaning* (Oxford University Press, 1998).

6. To claim that “ $x \text{ is true} \leftrightarrow x \text{ is } P$ ” is the explicit definition of the truth predicate (where “ $P$ ” might be replaced with “in correspondence with reality,” “verifiable,” “useful,” etc.) is to claim that our acceptance of such a principle is explanatorily fundamental with respect to our overall use of the truth predicate. But such a claim is incompatible with the minimalist thesis according to which it is rather our endorsement of the equivalence schema that is explanatorily basic.



7. However, such a theory of truth may be closely affiliated with the minimalist account of the meaning constitution of “true.” It may be that the axioms of the fundamental theory of truth—those that will provides the best explanation (i.e., simplest derivation) of all facts about truth—are instances of “ $\langle p \rangle$  is true  $\leftrightarrow p$ .” For (a) such axioms would suffice (in conjunction with theories of other matters) to explain every other fact about truth, and (b) it is hard to imagine a simpler body of principles on the basis of which those instances could themselves be explained. For further discussion, see my book *Truth*, pp. 25–31, 50–51 (1998).

8. D. Davidson, “The Folly of Trying to Define Truth,” *Journal of Philosophy* 87 (1996) 267–278. [See chap. 26.—Ed.] Objection 1, with its focus on compositionality, seems to me to present Davidson’s strongest reason for concluding, contrary to minimalism, that truth is conceptually prior to meaning. However, he does not elaborate this point in any detail, but instead stresses various other arguments for that conclusion. First, he infers it from the fact that there are sentences (such as “That is red”) whose meaning-constituting, assertibility conditions are to accept those sentences only when they are *true*. But this reasoning presupposes (wrongly, in my view) that meaning-constituting regularities of use are *explicitly* known by the speakers of a language. Second, he infers it from the fact that someone’s acceptance of a sentence may be justified only if he believes that the sentence is true. But that fact is better explained by minimalism, together with the fact that one should only assert what one believes. And third, he urges a direction of explanation which goes from facts about the circumstances that cause the acceptance of sentences, to facts about their truth conditions, to facts about their meanings. But there is a long-standing, notorious difficulty with that line of thought, namely, articulating a conception of truth conditions that is strong enough. For, given a *material* construal of the conditional, “*p*” may very well be true iff *q*, without its meaning that *q*. And stronger construals of “iff” merely make such counterexamples slightly harder to construct.

9. D. Davidson, “Truth and Meaning,” *Synthese* 17 (1967): 304–323.

10. For further discussion of the deflationary approach to compositionality, see my essay “The Composition of Meanings,” *Philosophical Review* 106 (1997): 503–531, reprinted as chapter 7 of my book *Meaning* (1998). In that work I suggest that the meaning of a complex expression is constituted by its being constructed in a certain way from words with certain meanings, and I offer this constitution thesis as an explanation of compositionality. But now it seems to me preferable to refrain from the constitution thesis and to take the principle of compositionality as explanatorily fundamental. For more detail, see my “Deflating Compositionality” (*Ratio*, forthcoming).

11. See my book *Meaning*. A rationale for the use theory of meaning is sketched at the beginning of this paper, where I motivate the minimalist view of how the meaning of the truth predicate is constituted.

12. One might be tempted by a more holistic theory in which truth and meaning (and proposition) are on the same conceptual level and are jointly explained in terms of one another and in terms of other matters. Such a theory would be less

simple and less explanatory, and hence less attractive, than a more atomistic view such as the one proposed here, and should be embraced only as a last resort.

13. D. Davidson, "Truth and Meaning" and "The Folly of Trying to Define Truth."

14. Davidson is careful to emphasize that his critique of minimalism does not amount to the claim that there are no such things as propositions. Presumably, he thinks that if propositions exist, they must be designated by expressions of the form "the proposition expressed by *u*" (where "*u*" refers to a sentence token), rather than by expressions of the form "the proposition that *p*." But even if this point were correct—and I have been arguing that it is not—the essence of the minimalist proposal would not be affected. For it could be reformulated in terms of the schema "The proposition expressed by the following sentence token is true  $\leftrightarrow p$ ." We could suppose that this is what fixes the meaning of "true."

In addition to Davidson's critique of minimalism, which focuses on how propositions are *designated*, there are several objections to the very *existence* of propositions—objections that a minimalist must be able to rebut. The main ones are (1) that propositions lack satisfactory identity conditions, (2) that *false* propositions do not exist (because any actual combination of objects and properties would amount to a *fact*, and (3) that propositions are ontologically weird and explanatorily unnecessary. For discussion of these issues, see my book *Truth* (1998).

15. H. Field, "Critical Notice: Paul Horwich's *Truth*," *Philosophy of Science* 59 (1992): 321–330.

16. A. Gupta, "A Critique of Deflationism," *Philosophical Topics* 21 (1993) 57–81 [see chap. 23.—Ed.]; and his "Minimalism," *Philosophical Perspectives* 7 (1993): 359–369.

17. A. Gupta, "A Critique of Deflationism" and "Minimalism."

18. A. Gupta, "A Critique of Deflationism" and "Minimalism." This objection has also been forcefully articulated by Scott Soames in his "Truth about Deflationism," in E. Villanueva (ed.), *Truth* (Atascadero, Calif.: Ridgeview Publishing Co., 1997). See also "Some Remarks on Deflationism" (unpublished) by Paolo Casalegno (University of Milan).

19. Although this strategy works for "All propositions of the form  $\langle p \rightarrow p \rangle$  are true," one might well wonder whether *all* general facts about truth can be explained in this way. But I think that we have some reason to think that they can be. For it would seem that any such fact could be put into the form "All propositions of type *K* have property *J*." Here are some examples: (1) Given any conjunction, if it is true, then so are its conjuncts. (2) Given any proposition of the form  $\langle p \rightarrow q \rangle$ , if it and its antecedent are both true, then so is its consequent. (3) Given any atomic proposition, it is true if and only if its predicate is true of the referent of its subject. And so on. Now, for any such generalization, if we can show, with the help of the equivalence schema, that it holds of an arbitrary proposition, we can invoke the proposed additional premise to explain our acceptance of that generalization.

Hartry Field, in his “Deflationist Views of Meaning and Content” (*Mind* 103, [1994]: 249–285), offers a solution to the present “generalization” problem that supposes the following, roughly speaking: (1) We can reason *schematically*, e.g., we can prove and assert “ $(p \ \& \ (p \rightarrow q)) \rightarrow q$ ,” and we then have the right to substitute sentences for schematic variables. (2) We can introduce a rule allowing us to go from any such schematic theorem, “ $\langle p \rangle$  is  $K$ ,” to the conclusion “All propositions are  $K$ .” (3) We in fact assert the schema “ $\langle p \rangle$  is true  $\leftrightarrow p$ ” (and not merely its instances).

My proposal differs from this merely in that, instead of (3), I have a claim (in effect) about the circumstances in which our disposition to believe all propositions of a certain form entitles us to assert the corresponding schema. This enables me to preserve the thesis of this paper—namely, that the meaning of “true” is fixed by our inclination to accept *instances* of the equivalence schema—rather than having to move to the claim that it is our inclination to accept *the schema itself* that fixes the meaning of the truth predicate. Asserting of a *schema* strikes me as too sophisticated an activity to be plausibly attributed to ordinary people.

20. A further objection of Anil Gupta’s—one that I *do* think is correct—is that our underived endorsement of the equivalence schema will not explain our confident acceptance of sentences like “Julius Caesar was not true.” To accommodate this point, we can suppose that the explanatorily basic, meaning-constituting facts about “true” include, not merely our underived allegiance to the equivalence schema, but also our underived acceptance of the principle that only propositions are true.

21. M. Richard, “Deflating Truth,” in E. Villanueva (ed.), *Truth* (1997).

22. I would argue that the moral of the ‘liar’ paradoxes is that not *all* instances of the equivalence schema are correct. But I don’t believe that those who come to accept this moral, and who come to balk at certain instances, are thereby altering what they mean by the truth predicate. This is my motivation for supposing that the meaning-constituting fact about “true” is a mere *inclination* to accept any instance of the schema, rather than a *disposition* to accept any instance. In problematic cases the inclination will be overridden. But its continued existence is what sustains the sense of paradox.

23. N.B. the division of a scientific theory,  $T(f_1, \dots, f_n)$ , into its Ramsey sentence,  $(\exists x_1) \dots (\exists x_n) T(x_1, \dots, x_n)$  (which says that there exist properties that relate to one another and to observable facts just as the theory says) and the conditional,

$(\exists x_1) \dots (\exists x_n) T(x_1, \dots, x_n) \Rightarrow T(f_1, \dots, f_n)$  (which says that *if* such there are such properties, they are  $f_1, f_2, \dots, f_n$ ).

It is plausible that just the second of these commitments is needed to fix the meanings of the theoretical terms. Consequently, interesting a priori knowledge cannot be supposed to derive merely from our understanding of words or our grasp of concepts. For further discussion, see my “Implicit Definition, Analytic Truth, and A Priori Knowledge” (*Noûs*, 1997; reprinted, slightly revised, as

chapter 6 of *Meaning*, 1998), and my “Stipulation, Meaning, and Apriority,” in *New Essays on the A Priori*, edited by P. Boghossian and C. Peacocke (Oxford University Press, 2000).

24. M. Dummett, “Truth,” *Proceedings of the Aristotelian Society* 59 (1958–1959): 141–162. [See chap. 10.—Ed.]

25. I would argue that the basis for our commitment to these specific norms is *pragmatic*: we are more likely to get what we want if we abide by them. See chapter 8 of *Meaning* (1998) and my “Norms of Truth and Meaning,” in *What Is Truth*, edited by Richard Schantz (Berlin and New York: de Gruyter, 2001).

26. Clearly, this is merely a first approximation of the proper norm. For one thing, there is nothing wrong with not bothering to investigate, and hence failing to believe, certain extremely trivial facts.



---

## The Metaphysics of Truth

Michael Devitt

### 1 Introduction

The most popular theory of truth has probably been the *correspondence* theory. According to this theory, the truth of a statement or belief consists in some sort of correspondence between the statement or belief and the world.<sup>1</sup> And truth is usually thought to play an important role in metaphysics. The correspondence theory has received several challenges.

1. Karl Popper and some logical positivists once thought that truth talk was unacceptably metaphysical and should be banished. They had the *eliminativist* view that statements are not, from a properly scientific perspective, really true or false at all. But, since Tarski at least,<sup>2</sup> this challenge to correspondence truth has disappeared.

2. Another challenge has come from the *verificationist* theory of truth, a theory that has received a boost from the work of Michael Dummett. On this view, the truth of a statement consists in its being warrantably assertable, verifiable, or something similarly epistemic. And this is thought to be metaphysically important because of its association with antirealism about the external world. I think that verificationist truth does indeed lead to this antirealism and that, from a naturalistic perspective, this is sufficient reason to reject it (Devitt 1997, secs. 4.3, 14.9). I shall not discuss it here.

3. In my view, the most interesting challenge has come from the *deflationary* theory of truth. The contrast between this influential theory and the correspondence theory has been much discussed. Yet, I shall argue, the contrast remains unclear. That unclarity arises from insufficient attention to the distinction between *the metaphysics of truth* and *the linguistics of the truth term*, and hence from insufficient attention to what the theories say, or should say, about the metaphysical issue. In arguing this in part I, I shall emphasize that deflationism is similar to a sort of “non-factualism.” Then, against this background, in the much briefer part II, I shall summarize the case for the correspondence theory.

## Part I Distinguishing the Deflationary Theory from the Correspondence Theory

### 2 Four Problems

What exactly is the difference between the deflationary theory and the correspondence theory? The answer is not as easy to find as one might have expected. There are four related problems in finding it.

The first problem is that the two theories have opposite focuses. Whereas the focus of the correspondence theory is on the nature and role of *truth*, the focus of the deflationary theory is on the nature and role of *the truth term*, for example, of ‘true’. The former focus is metaphysical; the latter, linguistic. So an awful lot of what deflationists say does not bear directly on what the correspondence theorists say, and vice versa.

A simple explanation of this difference in focus—too simple as we shall soon see—is as follows. Deflationism is really a sort of eliminativism, or antirealism, about truth: it deflates truth itself. We might say, *very* roughly, that according to deflationism, there is no reality to truth.<sup>3</sup> Since there is no reality to truth, there is nothing positive to be said about the nature of truth. However, unlike the eliminativists of challenge 1, deflationists have no objection to the use of the truth term. Indeed, they are enthusiastic about the term and have a great deal to say about its linguistic role and semantics.<sup>4</sup> In contrast, correspondence theorists are realists about truth and therefore struggle to explain its nature. But, for them, the truth term is just another one-place relational predicate—like, say, ‘warranted’ or ‘patriotic’—with the standard sort of semantics of such predicates. This semantics is likely to start from the assumption that the term denotes the property *truth* or applies to all true things. This is so unexciting as to be hardly worth saying, and the theorists are not usually inclined to say anything more exciting.

In sum, the deflationist has little to say about the metaphysics of truth but much to say about the linguistic role of ‘true’, whereas the correspondence theorist has a lot to say about the metaphysics of truth but little to say about the linguistics of ‘true’.

I am here describing a *real* difference in focus between the two theories. The second problem in distinguishing the theories is that this difference is often not apparent. Discussions of deflationism tend to blur

the distinction between the linguistic and the metaphysical. In particular, remarks that should be about the truth term are often presented as being about truth: there is use/mention sloppiness, even confusion. So it can often seem that discussions are talking about truth when they are really not.

As a result of the “linguistic turn” in philosophy, it has become common to slip casually back and forth between talking of words (or concepts or notions) and talking of the world. Probably this is harmless in most cases, but in some it is not. One of these is realism about the external world. Another is truth. But, in my view, whereas it is relatively easy to avoid confusion in the case of realism, it is rather hard to avoid it in the case of truth.

The third problem is that when discussions of deflationism do address the metaphysical issue, rather than merely appearing to when addressing the linguistic issue, what is said is often unsatisfactory. And this is not surprising, because it turns out to be rather hard to capture the deflationary metaphysics of truth. That is the fourth problem. A sign of this problem is that my characterization of the metaphysics of deflationism in describing the first problem *really is* very rough.

To appreciate the third and fourth problems, it helps to notice that deflationism is similar to “nonfactualism” in ways to be explained (section 4).

Despite these four problems, the difference between deflationary and correspondence theories on the linguistic issue of the truth term is relatively clear. Not so the difference on the metaphysical issue of truth. As a result of the four problems, there is a good deal of uncertainty, if not confusion, over the difference between deflationary and correspondence views of the nature of truth. This is serious because this metaphysical difference is deeper than the linguistic one, since it is explanatorily prior.

Section 3 will be concerned with the linguistics of the truth term. Sections 4 to 6 will be concerned with the deflationary metaphysics of truth and the third and fourth problems. I will criticize a standard characterization of the deflationary metaphysics and attempt a better one. These accounts of the linguistic and metaphysical issues are necessary background for appreciating the evidence in section 7 of the very tricky second problem: a use/mention sloppiness that obscures the real meta-



physics of deflationism. The uncertainties and confusions arising from these four problems will emerge as we go along but will be particularly prominent in section 7.

In talking of uncertainty and confusion, I speak from bitter experience, for I am only too well aware that my own writings on truth have provided some examples.<sup>5</sup>

### 3 The Truth Term

The deflationists have some very interesting things to say about the truth term. They have persuasively demonstrated that the term has an extremely useful “logical” or “expressive” role. Thus, suppose that Jack says, “We all lie about our sex lives,” and Jill replies, “That is true.” Intuitively, the role of the truth term here is to enable Jill to “say the same thing” as Jack without repeating his words (and while admitting that he said it first). Attention to such examples encourages the simplest deflationary theory, the “redundancy” theory, for they make it seem as if we could dispense with the truth term altogether.

However, other examples show that the term is very useful. It enables us briefly to assert something that may otherwise be tedious, if not impossible, to assert. Suppose that Imogen wishes to express general but qualified agreement with a certain article. She can say simply, “Most of what that article says is true.” Consider what would be required to say this without using ‘true’. Her claim entails that at least half of the claims in the article are true, but it is not specific about which claims. So her claim is equivalent to a long disjunction of conjuncts, each conjunct consisting of a different set of more than half the claims in the article. If she could remember all the claims, she could, in time, manage to express this disjunction. If not, she needs the truth term. So does a person who has forgotten Goldbach’s Conjecture but nevertheless wants to express agreement with it. He can say, “Goldbach’s Conjecture is true.” A person who has lost track of all the utterances of the Great Helmsman can nevertheless express her commitment, ‘Everything Chairman Mao said was true’. Without the truth term, she faces the impossible task of asserting an infinite conjunction. So also does a logician in asserting each instance of a schema that has an infinite number of instances.<sup>6</sup>

The truth term can play its logical role because it yields equivalences like the classic one between “‘Snow is white’ is true” and ‘Snow is white’. When the term is attached to the quotation name of a statement, it yields a statement that is equivalent to that statement: it undoes the effect of quotation marks. (Attention to this led to the name ‘the disquotational theory of truth’.) Indeed, when the truth term is attached to any device for referring to a statement, it yields a similar equivalence; it is a “denominalizing” device. Thus Jill’s remark, “That is true,” is equivalent to Jack’s, “We all lie about our sex lives.” If I were to say, “Jack’s remark is true,” my remark would be equivalent to Jack’s. The person who said, “Goldbach’s Conjecture is true,” said something equivalent to the Conjecture. In general, the deflationary view supports “the equivalence thesis”: all appropriate instances of the “equivalence schema”(1) hold, where an appropriate instance substitutes for ‘*p*’ a translation of the statement referred to by what is substituted for ‘*s*’:<sup>7</sup>

(1) *s* is true iff *p*

What is the “meaning” of the truth term? The deflationists have offered a variety of answers. Thus, Paul Horwich (1990) proposes a “minimal” theory according to which ‘true’ is an unusual “logical” predicate implicitly defined by its use in the appropriate instances of the equivalence schema. Dorothy Grover (1992) urges a “prosentential” theory according to which ‘true’ is not a predicate at all. Rather it is a syncategorematic part of an anaphoric “prosentence,” where prosentences are to sentences as pronouns are to nouns. This ingenious theory has the unhappy consequence that ‘that’ in ‘that is true’ does not *refer* to some statement, as one would naturally suppose (and as I supposed in introducing the equivalence thesis). This led Robert Brandom to propose a variation on the prosentential theory that avoided this consequence: the truth term should be treated as a prosentence forming operator (1988, 88–90).

So much for deflationary views of the meaning and role of the truth term. What are correspondence theorists to make of this? The important thing to notice is that they can, and should, go along with most it. Certainly, they cannot go along with a deflationary theory of meaning of

the truth term, whether of the Horwich, Grover, Brandom, or any other variety. They think that the term has the standard semantics of a one-place relational predicate, very likely explained in terms of reference to the truth property or to true statements, as I noted. Still, they can and should accommodate the rest of the deflationary story. In particular, they should accept the equivalence thesis, for that is a constraint on *any* theory of the truth term. And if the correspondence theory meets that constraint, it can account for the logical role of ‘true’ that the deflationists have so persuasively demonstrated.<sup>8</sup>

So although the correspondence theorist disagrees with the deflationist over the meaning of the truth term, he should agree that the term has the logical role explained by the deflationist. There will probably be one other important disagreement. The deflationist will insist that the truth term does not have any role other than the logical one; in particular it does not have the “descriptive” role of a normal predicate. The correspondence theorist is likely to think that the term has a substantial descriptive role in some theory of the world.

These linguistic differences between a deflationary theory and a correspondence theory over the truth term are striking and obvious. The metaphysical differences between the two theories over the nature of truth are much less so. Yet the metaphysical differences are explanatorily prior because they largely motivate the linguistic ones.

#### 4 The Deflationary Metaphysics of Truth: The Problem

We have seen what the deflationists say about the truth term, but what is their view of truth? Where do they stand on the metaphysical issue? I have said that deflationism is a sort of eliminativism or antirealism, and I roughly characterized it as denying that there is any reality to truth (section 2). But the inadequacy of this is apparent when we note that deflationists are as ready to talk about statements being true as correspondence theorists; thus many deflationists will say that Jack’s remark is true, because to say this is just to express the common belief that we all lie about our sex lives; and they will all agree that ‘Snow is white’ is true, because to say this is just to say that snow is white. So what does their antirealism consist in? The focus of deflationist literature is not on answering this question, and the little that the literature says is often

unsatisfactory. This was our third problem in section 2. The question also turns out to be rather hard to answer. This was our fourth problem.

To appreciate these two problems it helps to realize that analogous problems arise elsewhere. For deflationism about truth is similar to the “nonfactualism,” exemplified by “noncognitivism,” about morals, “projectivism” about causality, positivistic instrumentalism about science, and Simon Blackburn’s “quasi realism” (1984, 1993a, 1993b).<sup>9</sup> Characterizations of the metaphysics of nonfactualism also tend to be unsatisfactory, and it is difficult to give a satisfactory characterization.

I have discussed these problems for nonfactualism elsewhere and will draw on those discussions in what follows.<sup>10</sup>

Deflationism has two defining features of this kind of nonfactualism. The first is at the linguistic level and is very explicit in the literature. Nonfactualism in an area has a revisionist view of the language in that area: the language is not “descriptive,” as we would naturally take it to be. This view is expressed in a variety of ways, some rather unsatisfactory, but the key idea is clear: terms that appear to be predicates in the area do not have the standard semantics of a normal predicate; perhaps they are not predicates at all. Because these terms are in this way “non-descriptive,” they are not like a normal predicate in purporting to “describe reality”; they have some other role. Thus the most famous nonfactualism, noncognitivism about morals, has a revisionist view of the semantics of ‘good’ as a result of which indicative sentences containing it are not assertions or statements. Rather, those sentences express attitudes or emotions, or prescribe norms or rules.<sup>11</sup> The deflationist view of the truth term, discussed in the last section, is a similar sort of revisionism. The truth term does not have the standard semantics of a normal predicate. And its role is not to describe sentences; its only role is logical or expressive.

What is meant by “the standard semantics”? Typically, a philosopher’s standard semantics will be truth-referential, but it need not be: it might be verificationist, for example. And the standard semantics of a deflationist cannot be truth-referential, because, for her, truth is not explanatory (section 5). So the non-truth-referential meaning that she attributes to a normal predicate—for example, a certain sort of use condition—she does not attribute to the truth term.

Despite the linguistic similarities between deflationism and nonfactualism, there are important differences that should make one reluctant to treat deflationism as a species of nonfactualism. First, the deflationist's "expressive" role for 'true' is nothing like the noncognitivist's "expressive" role for 'good': the former is logical; the latter emotive. Second, the noncognitivist holds that because 'good' is nondescriptive, sentences of the form '*x* is good' are not factual. In contrast, the deflationist does not hold that because 'true' is nondescriptive, sentences of the form '*S* is true' are not factual. For her, whether these sentences are factual depends on whether *S* is factual. So if *S* is '*x* is good', it is factual for a deflationist who is a cognitivist but not factual for one who is a noncognitivist.

The second defining feature of nonfactualism is at the metaphysical level and is often more implicit than explicit in the literature. Nonfactualism in an area is antirealist about that area. Thus noncognitivists are antirealist about goodness. Deflationists are similarly antirealist about truth.

Consider the problem of characterizing the antirealism of nonfactualism. The most straightforward way of characterizing antirealism in general, using the ordinary language for denying ontic commitment, obviously does not capture the metaphysics of nonfactualism. Thus the noncognitivist does not claim that there are no good people, right actions, and so on. She is as ready as the realist to say, "This person is good" and "That action is right," for she thinks that these utterances express appropriate emotions or prescriptions. We have already made the analogous point about deflationism: the deflationist does not claim that there are no true statements. The nonfactualist and the deflationist talk like a realist but give that talk a revisionist interpretation.<sup>12</sup> This is what poses the problem of distinguishing these doctrines from realism at the metaphysical level (our fourth problem).

But perhaps the problem is illusory. Maybe my confident claim that nonfactualism and deflationism are antirealist is mistaken. Perhaps the focus of these doctrines is so linguistic because they really have no commitment one way or the other on the metaphysical issue. If this were so, our enterprise of attempting to characterize their antirealism would be misguided. There are two reasons why this dissolution of the problem must be rejected. The first is that these doctrines are presented in oppo-

sition to realist views; thus, deflationists oppose correspondence truth. And despite the linguistic focus, the doctrines are accompanied by claims that are clearly intended to be antirealist even if, as we shall see, the claims are often not adequate to the intention. The second reason is that an antirealist metaphysics is needed to *motivate* the revisionist view of language urged by these doctrines. If there were not something problematic or defective about the area of reality that 'true' or 'good' appear to concern, why suppose that they do not have the standard semantics of a descriptive predicate? Of course, the semantic revisionism is typically supported by some purely linguistic considerations: evidence of a non-descriptive role for the language in question. Thus deflationists are motivated by the logical role of the truth term, and noncognitivists by the action-guiding role of moral language. But what is to stop language covered by the standard semantics from playing these roles? Indeed, we have already suggested that a truth term with the standard semantics could play the logical role (section 3). So the antirealist metaphysics is still needed to make the standard semantics for this language unattractive. It is needed to show that the language does not have a descriptive role *as well as* the role emphasized by nonfactualism and deflationism. Behind the linguistic facade of these doctrines must lie an antirealist metaphysics.

Because the metaphysics of deflationism is needed to motivate its semantics, the metaphysical difference between deflationism and the correspondence theory is explanatorily prior to the linguistic difference.<sup>13</sup> So the problem of characterizing the antirealism of nonfactualism and deflationism remains.

The most straightforward characterization is obviously hopeless and may have no adherents. However, another simple characterization is popular: the doctrines are said to deny that there are any *properties* in the area in question.<sup>14</sup> Thus noncognitivism denies that there is a property of *goodness*, and deflationism that there is one of *truth*.

As soon as we look carefully at this popular characterization, we see that it cannot be satisfactory. This is our third problem with deflationism, and there is an analogous problem with nonfactualism. The characterization is unsatisfactory because it overlooks the extent to which a philosopher's attitude to the metaphysics characterized might reflect a position

on the *general* issue of realism about properties rather than a position on the particular problematic area of reality that is the concern of the non-factualist or deflationist, for example, rather than a position on morality or truth. Thus, consider a *nominalist*. She will agree that there are no properties of goodness and truth because she thinks that there are no properties at all! Yet, manifestly, this alone does not commit her to non-factualism and deflationism; to thinking that there is something especially defective about the realms of morals and truth, something that motivates a revisionist semantics. She might be as realist as could be *about morality and truth*. Or consider someone like David Armstrong (1978), who is a *selective realist* about properties. Armstrong thinks that empty predicates, disjunctive predicates, and negative predicates have no corresponding properties. He thinks that some predicates apply to the world in virtue of many properties. Most important, he looks to science to tell us which properties there are. Such a person might well be a reductive realist, thinking that ‘good’ or ‘true’ apply to an object in virtue of properties *none of which are goodness or truth*; they apply in virtue of scientifically acceptable properties. So the popular characterization fits his views even though his metaphysics of goodness and truth is quite contrary to the antirealist one that we are attempting to characterize. Finally, consider the *unselective realist*, who thinks that there is a property for each predicate. A nonfactualist might accept that ‘good’ is a predicate, as indeed Blackburn (1993a, 206) does, and a deflationist might accept that ‘true’ is, as indeed Horwich (1990) does. If such a person is an unselective realist, she will think that there *is* a property of goodness or truth, thus *disagreeing* with the popular characterization. And even if the non-factualist denies that ‘good’ is a predicate, and the deflationist that ‘true’ is, and hence that there are properties of goodness and truth, the popular characterization of their antirealism is dubious: it runs the wrong way. It finds a defect in reality because of something special about language, whereas we need to find a defect in reality to motivate the view that the language is special.<sup>15</sup>

The general issue of realism about properties is independent of the issues of nonfactualism and deflationism. It should be possible for someone to embrace or reject the metaphysics of these doctrines whatever her position on this general issue. There should be a way of stating that

metaphysics that is appropriate whatever the truth of the matter about the reality of properties.

So far, then, we have made no progress characterizing the antirealism of nonfactualism and deflationism. The most straightforward statements of realism, using the ordinary language of ontic commitment, are not denied by these doctrines, because they are reinterpreted so that they have no such commitment. We have just seen the failure of a characterization using more “philosophical” talk of properties. In general, the nonfactualist/deflationist practice of talking like a realist while giving that talk a revisionist interpretation makes progress hard. We are attempting a characterization of the metaphysics that must motivate the special semantic treatment that the doctrines give to a certain area of language. Yet our attempts seem vitiated by that very semantic treatment. Nonfactualism and deflationism are supposed to be a sort of antirealism, and yet it seems impossible to give a metaphysical statement of their antirealism. Issues of realism begin to evaporate. Indeed, Blackburn sometimes comes very close to claiming that they have evaporated (1993a, 4, 15–34, 55–59; 1993b, 368).

## 5 The Deflationary Metaphysics of Truth: The Solution

To avoid the evaporation of realism as a metaphysical issue and to characterize the metaphysics of nonfactualism in an area, we must first find some language in that area that is not just apparently descriptive but is treated by the nonfactualist as *really* descriptive. We must then examine her statements using that language to find ones that disagree with realist statements about the area.

I have argued elsewhere (1996b, 165–170; 1997, 313–318) that two sorts of realist claim are the most promising candidates for denial by the nonfactualist. First, the typical realist offers explanations of the *nature* of the problematic reality in language that the nonfactualist should agree is factual. For the realist thinks that the problematic reality is constituted by, or supervenes on, a reality that should be *unproblematic* for the nonfactualist. Even though the nonfactualist claims to be able to accept many sentences that seem to describe the problematic reality by taking them as expressive, prescriptive, or whatever, she does not accept the need for, or possibility of, these substantial “broadly reductive” expla-



nations. Thus moral realists claim that there are things about a person in virtue of which she is good, that make her good; for example, being kind, considerate, generous, honest, etc. And there are things about an action in virtue of which it is wrong, that make it wrong; for example, leading to unhappiness, being contrary to socially accepted rules, and so on. The noncognitivist must reject all such “in virtue of” claims as totally misconceived.

The deflationist has a similar disagreement with the typical realist about truth. The realist will claim that there is something *common and peculiar* to true statements: a statement is true *in virtue of* some sort of correspondence relation to the world; this relation *makes* it true. A substantial theory is then needed to describe and explain this correspondence, a theory that may include, for example, causal theories of reference. Deflationists should reject any such reductive explanation of truth. Horwich does so in denying that truth has an “underlying nature” or some “hidden structure awaiting our discovery” (1990, 2): “*Being true* is insusceptible to ... scientific analysis” (1990, 6). Grover claims that “truth talk ... can be explained without appeal to any kind of analysis of the nature of truth” (1992, 3)

This is not to say that the deflationist rejects all statements of the form ‘*p* explains that *S* is true’. The deflationist, like everyone else, accepts the need for, and possibility of, explanations of “worldly facts” such as that snow is white, explanations that appeal to laws of nature. Suppose that *E* explains that snow is white. So, given the deflationary theory, *E* explains that ‘Snow is white’ is true. But this sort of explanation, varying from truth to truth, is not what the correspondence theorist seeks. He seeks an explanation of what all true statements have in common, an account of “correspondence to the world.” That is the sort of explanation that the deflationist must reject.

The second sort of realist claim that the nonfactualist should deny concerns *causal role*. The typical realist thinks that the problematic reality is the cause or effect of some unproblematic reality. The nonfactualist should not accept these claims about the role of the problematic reality, because on her view there is no reality that could play such a role. Thus the typical moral realist thinks that there are causes and effects of a person being good or bad. He thinks that it is because Hitler and his associates

were depraved that we believe that they were depraved. And it is because they were depraved that they behaved as they did and that millions of people died in concentration camps. The noncognitivist must reject all such explanations.

Once again, the deflationist has a similar disagreement with the realist about truth. The typical realist will give truth important explanatory roles, for example, to explain the success of science or the success of people in meeting their goals, or to explain meaning, where meaning itself plays a role in the explanation of behavior. A deflationist must reject all such explanations, and Brandom clearly does reject them all (1988, 91–92).<sup>16</sup>

This is not to say that the deflationist cannot use the truth term in explanatory statements. For the deflationist, the logical role of the truth term makes an explanation of the form '*p* because it is true that *q*' equivalent to one of the form '*p* because *q*'. But the appearance of 'true' in the former sort of explanation does not make truth explanatory of *p*. Consider an example: 'Clinton was impeached because he is hated by the religious right' can be rewritten as 'Clinton was impeached because it is true that he is hated by the religious right'. Manifestly, what is explanatory here is hatred, not truth. Even where the expressibility provided by the truth predicate is essential to an explanation—because without it the explanation would be infinite—it is not truth that is explanatory.

In sum, the typical realist thinks that there is a reality to truth that, like any other reality, has a nature and causal role, and that this nature and role need explanations. The deflationist reveals her antirealism by rejecting the need for, and possibility of, such explanations. Although she can join with the realist in accepting ordinary truth claims, she cannot join with him in his explanation of the reality which he takes those claims to describe. The deflationist should have nothing that is positive and substantial to say about truth.

Sadly, this account of the distinction between realism and nonfactualism/deflationism has a flaw, reflected in my frequent uses of 'typical'. There are doubtless some philosophers who claim to be moral realists and yet join the noncognitivists in denying the need for an explanation of moral reality and in denying that this reality has any causal role: it is inexplicable and epiphenomenal. One can imagine an analogous claim from someone who sees himself as a realist about truth. Such positions are deeply anti-

naturalist, of course. They are also hard to motivate: why believe in a truth or goodness that can do nothing and cannot be explained? Still, the positions are possible. And if they have a standard semantics for 'true' and 'good', they surely are realist, for they accept the straightforward statements of realism without interpreting away the ontic commitment of the statements. So the flaw in my account is that it does not distinguish this atypical realism from nonfactualism and deflationism at the strictly metaphysical level. I suspect that this realism cannot be so distinguished. If not, we must conclude, disappointingly, that to *fully* capture the anti-realism of nonfactualism/deflationism, we have to add a little semantics: what makes these doctrines antirealist is not only their denial of explicable nature and causal role but also their adoption of a nonstandard semantics that removes the commitment from apparently straightforward statements of realism.

The nonfactualist/deflationist and the atypical realist agree that in a certain area there is no reality with an explicable nature and a causal role. Despite this failure, the atypical realist holds that there is a reality in that area: the reality is simply inexplicable and epiphenomenal. The failure motivates the nonfactualist/deflationist, in contrast, to reject the reality altogether by revising the semantics for what would otherwise be straightforward statements of realism.

## 6 The Equivalence Thesis

The difference between deflationism and the correspondence theory should emerge in proponents' responses to a demand for an explanation of the equivalence thesis. Let us take the most famous instance of the equivalence schema as our example: 'Snow is white' is true iff snow is white. In virtue of what is this so?<sup>17</sup> The core of the correspondence theorist's answer is a reductive theory of the nature of true statements. This will be an account of what is common and peculiar to these statements and of the relation that true statements stand in to the world. When this theory is applied to 'Snow is white', it shows that this statement is related to the world in such a way that the statement is true iff snow is white. So the theory of truth, together with facts about the statement 'Snow is white', explain why 'Snow is white' is true iff snow is white.

The deflationist, in contrast, cannot accept any appeal to a theory of the nature of truth in her explanation, because she dismisses the possibility of saying anything substantial about that nature. So what explanation does she offer? Basically, none. She thinks that the demand for an explanation here is misguided: that ‘Snow is white’ is true iff snow is white is a “brute fact” needing no explanation. However, she has something further to say to make this provocative claim palatable: a diagnosis of the error of thinking that we need an explanation here. The diagnosis moves up to “the semantic level” to consider how the brute fact is expressed. Although the deflationist denies the need to explain why ‘Snow is white’ is true iff snow is white, she accepts the need to explain why people wrongly think that the statement “‘Snow is white’ is true iff snow is white” expresses something that needs explaining. The error arises from treating ‘true’ as if it were a normal descriptive relational predicate, thus taking the truth of ‘Snow is white’ to depend on some relation that this statement has to snow’s being white. Once the nondescriptive meaning of ‘true’ is appreciated, we see that to say that ‘Snow is white’ is true is not to relate the statement in some way to the world but simply to say that snow is white. So, *of course* ‘Snow is white’ is true iff snow is white, just as snow is white iff snow is white.<sup>18</sup> No more needs to be said (unless “the logical structure of the world” is to be explained). Similarly, it might be claimed that the following are brute facts needing no explanation: that Schnee ist weiss iff snow is white, that all bachelors are unmarried, and that Hesperus is Phosphorus. However, someone might think otherwise because she failed to appreciate the relevant semantic facts: that ‘Schnee ist weiss’ is synonymous with ‘Snow is white’, that the meaning of ‘bachelor’ includes the meaning of ‘unmarried’, and that ‘Hesperus’ and ‘Phosphorus’ rigidly designate the same object.

The contrast between the two theories should not be that the correspondence theory must offer a substantial explanation of why ‘Snow is white’ is true iff snow is white where the deflationary theory offers a trivial one appealing to the meaning of ‘true’.<sup>19</sup> The contrast should be that the correspondence theory must offer an explanation where the deflationary theory appeals to the meaning of ‘true’ to explain why no explanation is necessary.

The position I am attributing to the deflationist on this matter is undoubtedly hard to grasp. The position is developed and modified a little in the next section.

I started this part of the paper by mentioning four related problems in distinguishing deflationism from the correspondence theory. The first problem was a difference in focus: the focus of deflationism is on the linguistics of the truth term; the focus of the correspondence theory on the metaphysics of truth. The third problem was the unsatisfactory nature of attempts to characterize the metaphysics of deflationism, and the fourth was the difficulty of such a characterization. I have said a lot about these three but nothing yet about the second problem. We now have the background to discuss it.

## 7 Use/Mention Sloppiness

The second problem was use/mention sloppiness, even confusion, in the literature: deflationist remarks that should concern the linguistics of the truth term are often misrepresented as being about the metaphysics of truth, thus obscuring the real metaphysics of deflationism. In giving examples of this sloppiness, I do not mean to suggest that all of them amount to real confusions in thinking. Some surely are just insignificant carelessness or convenient rhetoric. Still, I want to show, first, how pervasive the sloppiness is. Second, I want to make it plausible that there are some cases of real confusion: that what should be a theory of the truth term is really being taken as a theory of truth, not simply carelessly expressed. This shows, it seems to me, how very difficult it is to handle the use/mention distinction in discussing truth.

In the light of our discussion so far, it is easy to spot the sloppiness. The deflationist is talking about truth itself, and saying something appropriate, when she denies that truth has a nature or causal role that needs or can have an explanation (section 5). And she is talking about truth itself, but saying something inappropriate, when she denies that truth is a property (section 4). Anything else she says, particularly anything positive, that is represented as being about truth should very likely be about the truth term. Here, then, are six cases of use/mention sloppiness.

1 The problem starts with the very names of some deflationary theories. The name ‘the redundancy theory of truth’ implies that truth is redundant, yet what is really redundant according to the theory is the truth term. Similarly, what is really disquotational according to ‘the disquotational theory of truth’ is the truth term, not truth. What is prosentential according to the ‘the prosentential theory of truth’ is not truth but a linguistic expression including the truth term. The generic name ‘the deflationary theory of truth’ does refer to theories that deflate truth, not the truth term, and so the name does not confuse use and mention. Still, the name is a bit misleading because only a small part of what deflationary theories actually say concerns truth. What they say mostly concerns the truth term. They deny a descriptive role for the term but emphasize other roles that were largely unnoticed or ignored by correspondence theories. On balance, deflationary theories inflate the truth term.

2 Consider next a historically important but notoriously difficult case of deflationism: Alfred Tarski. A special difficulty is that Tarski sees himself not as a deflationist but rather, it seems, as a correspondence theorist (1956: 153, 404).

On the opening page of “The Concept of Truth in Formalized Languages,” Tarski variously describes his enterprise as the definition of *truth*, of the term ‘*true sentence*’, and of the *concept of truth* (1956, 152). The last two can be taken to be the same, but *prima facie*, they are different from the first.<sup>20</sup> Defining truth is a matter of explaining its nature, a metaphysical matter, whereas defining the term and the concept are linguistic matters.<sup>21</sup> We have use/mention sloppiness.

What does Tarski actually do? He defines the meaning of ‘true-in-*L*’, where *L* is any of a certain range of formal languages. Does this have anything to do with explaining the nature of truth? Set aside worries arising from the fact that he has defined ‘true-in-*L*’, not ‘true’, and suppose that he had defined ‘true’. Could that have shown anything about truth? It depends on the definition. In certain cases we can move straight from a definition of a word’s meaning to an explanation of the nature of the reality that the word concerns. For example, we can move straight from defining ‘bachelor’ as ‘adult unmarried male’ to the explanation

that to be a bachelor is to be an adult unmarried male. So moving back and forth between talk of defining 'bachelor' and defining bachelorhood would be an insignificant use/mention sloppiness, of interest only to pedants. But a linguistic definition licenses this move to metaphysics only if it treats the word in question as a normal descriptive predicate. Where the definition amounts to a revisionist view of the word's meaning, the definition cannot yield a substantial explanation of the nature of the reality that the word appears to concern.<sup>22</sup> Indeed, an antirealist view of that reality is necessary to motivate the revisionist semantics (section 4). Consider an example: we could not move from a noncognitivist definition of 'x is good' as 'Hoorah for x!' to an explanation of goodness as "hurrahness," and noncognitivism is partly motivated by an antirealist view of goodness. One lesson I think that we should draw from Hartry Field's classic article "Tarski's Theory of Truth" (1972) is that Tarski's definition of 'true' is of the revisionist sort and so, as it stands, does not show us anything substantial about truth. Tarski's use/mention sloppiness is of more than pedantic interest.

Tarski's definition of 'true-in-*L*' rests on listlike definitions of various referential words along the lines of the following definition of 'designate':

'N designates *x*' =<sub>df</sub>

'N is "France" and *x* is France or

N is "Germany" and *x* is Germany' or

⋮

By comparing such definitions with a similar one for 'valence', Field brings out dramatically that the definitions do not yield satisfactory reductive explanations of the nature of reference.<sup>23</sup> So the definition of 'true' in terms of the referential words does not yield a satisfactory reductive explanation of the nature of truth. In the light of subsequent discussions, we can see why: the listlike definitions are essentially deflationary,<sup>24</sup> and so *could not* yield anything substantial about reference. Indeed, in offering these definitions, Tarski is implicitly committed to antirealism about reference: only if there were something problematic about reference would there be adequate justification for not treating the referential terms as ordinary two-place relational predicates; for not

saying, for example, that ‘designate’ designates the relation *designation*, or applies pair-wise to all ordered pairs where the first member designates the second. Tarski shared the physicalism of the positivists and clearly did think that there was something problematic about both reference and truth. And that was the thought that drove his enterprise. Although Tarski seemed to view himself as a correspondence theorist about truth, the theory he actually presented is deflationary, as I think is now generally agreed. So there is a far from innocent use/mention confusion in representing Tarski’s definition as a theory of truth, as Tarski and others do.<sup>25</sup> Tarski’s definition tells us a lot about ‘true-in-*L*’. It tells us nothing about truth-in-*L*, because it is implicitly committed to the view that there is nothing to tell.<sup>26</sup>

3 Scott Soames begins “What Is a Theory of Truth?” (1984), an important defense of Tarski from Field’s criticisms, with the report, “Alfred Tarski’s theory of truth and its successors ... are commonly believed by philosophers to provide analyses of the nature of truth” (1984, 411). If Soames is right that this belief is common—and I think he is—the use/mention confusion I have just noted is widespread. Soames does not share my view that this belief misrepresents Tarski’s achievement, but he notes that “there is considerable doubt about whether, or in what sense, [Tarski’s theory] is a theory of truth.” He goes on,

One main reason for this uncertainty is the difficulty of determining what a theory of truth ought to be. Generally, theories of truth have tried to do one or the other of three main things:

- i. to give the meaning of natural-language truth predicates;
- ii. to replace such predicates with substitutes, often formerly defined, designed to further some reductionist program; or
- iii. to use some antecedently understood notion of truth for broader philosophical purposes. (1984, 411)

This is striking. Suppose that we wondered what a theory of, say, genes tries to do. Two things occur: (a) it tries to describe the role of genes—state the laws about genes—which is what Mendelian genetics does; (b) it tries to say what genes *are*—explain their nature—which is what molecular genetics does. Now explaining the role of genes is, near enough, analogous to Soames’s task (iii). But explaining what genes are has no



analogue on Soames's list! The metaphysical task of explaining what truth is, which is surely what correspondence theorists and many others were trying to do, has become one or another of the two linguistic tasks, (i) and (ii). Use has become mention.

4 The theory that Horwich proposes in his influential book *Truth* (1990) is explicitly deflationist. Yet he talks positively of "the minimalist function" of truth (1990, xii), of "the entire conceptual and theoretical role of truth" (1990, 6), of "the properties of truth" (1990, 26), and of "all the facts involving truth" (1990, 7). Strictly speaking, on his anti-realist theory, truth can have no function, role, or (nontrivial) property, and it cannot be involved in any facts. The truth term is what has the function, role, properties and involvement. Horwich claims that his minimal "theory of truth ... involves nothing more than the equivalence schema" (1990, 12); it is "what is expressed by [the schema's] uncontroversial instances" (1990, 7).<sup>27</sup> But this is misleading at best, for his theory of truth is really to be found in various negative remarks about the nature of truth, some of which I have quoted (section 5). Of course, the word 'true' is used, not mentioned, in each instance of the equivalence schema, and this might suggest that these instances explain the nature of truth. But that suggestion treats 'true' as a normal descriptive predicate, which is precisely what Horwich, like all deflationists, denies: he has a revisionist theory of the truth term.<sup>28</sup> This theory of the truth term is what really "involves nothing more than the equivalence schema": it holds that every fact about the role of the term can be explained simply by taking the meaning of the term to be implicitly defined by its use in the appropriate instances of the equivalence schema.<sup>29</sup> The equivalence schema has nothing to do with his theory of truth and everything to do with his theory of the truth term.<sup>30</sup>

An analogy with goodness may help. Suppose that a noncognitivist were to talk positively of the function, role, and properties of goodness, and of the facts involving goodness. Suppose that she claimed to be giving a theory of goodness that was quite clearly based on her views about the nondescriptive meaning and expressive role of 'good'; for example, she claimed that her theory of goodness involves nothing more than the view that to say that  $x$  is good is just to express a pro attitude toward  $x$ .

It would be obvious that she was misdescribing her position: on her view, such remarks should really apply only to 'good', not goodness. For her theory of goodness is, roughly, that there isn't any; more precisely, it is the view suggested in section 5.

5 Stephen Leeds's "Theories of Reference and Truth" (1978), and Robert Brandom's "Pragmatism, Phenomenalism, and Truth Talk" (1988) are two of the best brief presentations of deflationary truth.<sup>31</sup> Leeds sketches a disquotational theory of the sort famously suggested by Quine (1970). He comments, "What we have sketched is not a theory of truth ... but a theory of the *concept* of truth" (1978, 122). But then he spoils this assessment by claiming that his account explains "facts about truth-in-English" and "what we ordinarily say about truth" (Leeds 1978, 123). It doesn't. But it might explain facts about 'true' in English and ordinary uses of 'true'. Brandom takes the "central theoretical focus" of deflationism to be "on what one is doing when one takes something to be true, that is, our *use* of 'true'." He goes on: "It is then denied that there is more to the phenomenon of truth than the proprieties of such takings" (Brandom 1988, 77). But strictly speaking, on the deflationary view, the proprieties are not any part of the phenomenon of truth, because, roughly, there is no such phenomenon. The only phenomena are truth takings.

6 Finally, consider Marian David's *Correspondence and Disquotation* (1994), the most detailed and informed critique of the disquotational theory of truth available. David starts his description of the disquotational theory by claiming that it, unlike, say, the correspondence theory, is "an antitheory of truth": its view is, "Truth has no nature." So far, so good. But then he continues: "The correct explanation of truth ... requires less extravagant resources." The correct explanation is that "truth is disquotation" (1994, 3–4). But the disquotational view does not require *less* extravagant resources to explain truth, it does not require *any*, because, properly understood, it is the view that truth does not need and cannot have an explanation. That is the respect in which it "has no nature." And disquotation does not explain truth, it explains the truth term.

Not surprisingly, when David sets out to find the unextravagant disquotational theory of what it is for a sentence to be true, he finds the theory “a bit elusive” (1994, 62). The core of the disquotational theory is, of course, the equivalence schema. David worries away at the schema, trying unsuccessfully to find in it a theory of truth other than a correspondence theory. Sometimes he comes close to realizing that he is seeking something that the disquotationalists think is not there to be found, for they think that “sentence-truth is in a sense ‘nothing’” (1994, 65);

Strictly speaking, the question [about truth] will not even receive a response with the right logical form to count as an answer to this question, for the grammatical truth predicate does not function like an ordinary predicate.... Given that the standard way of answering “What is F?”—questions does not work when it comes to truth all one can do is describe the *linguistic role* that the term ‘true’ plays in our language. (1994, 68–69)

Just so. Still he remains puzzled: “where does the deflationary idea that truth is *nothing but* disquotation come from?” (1994, 69). Deflationists have given him reason to be puzzled, as we have seen. Despite what is often suggested, the disquotational view should not be that truth is nothing but disquotation. The view should be that truth is nothing.

In this section, I have indicated how pervasive use/mention sloppiness is in discussions of deflationary truth. Some of this sloppiness is surely insignificant. Yet I hope to have shown that some of it is not: a theory of the truth term is really being taken as a theory of truth. This helps to obscure the metaphysics of deflationary truth and hence the difference between deflationism and the correspondence theory described in sections 5 and 6.

## 8 Summary

In this part I have attempted to bring out the real difference between deflationism and the correspondence theory by emphasizing the similarity between deflationism and nonfactualism. At the linguistic level, the real difference is fairly apparent. The correspondence theorist can, and should, grant that the truth term has the logical role emphasized by the deflationist. But the correspondence theorist does not accept the deflationary view that the term has no other role: he holds that it also has a

descriptive role. Furthermore, he thinks that the term has the standard semantics of a one-place descriptive predicate—a view that the deflationist rejects. At the metaphysical level, the real difference between the two theories is much harder to discern. The typical correspondence theorist thinks that truth has a nature and causal role that need explaining. The deflationist should reveal her antirealism in the characteristic non-factualist way by rejecting the need for and possibility of any such explanation. Finally, the metaphysical difference implicitly, if not explicitly, motivates the linguistic difference: it is largely because of her antirealism that the deflationist rejects a standard semantics and a descriptive role for the truth term.

I have located the difficulty in discerning the metaphysical difference in four problems. The first problem is a difference in focus: the focus of deflationism is on the linguistics of the truth term; the focus of the correspondence theory is on the metaphysics of truth. The second problem, just illustrated in some detail, is that use/mention sloppiness in discussions of deflationism tends to obscure the real metaphysics of deflationism. The third problem is that when discussions do address the metaphysical issue, rather than merely appearing to when addressing the linguistic issue, what is said is often unsatisfactory. And this is not surprising, because it turns out to be rather hard to capture the deflationary metaphysics of truth, as it is to capture the metaphysics of nonfactualism. That is the fourth problem.

Having clarified the difference between the deflationary theory and the correspondence theory, I shall now summarize the case for the correspondence theory.

## **Part II The Case for the Correspondence Theory**

### **9 Knowledge of Equivalences**

I start my case by acknowledging something that may seem to count against the correspondence theory and in favor of the deflationary theory. It is a striking fact that people competent with the truth term tend to believe instances of the equivalence schema, for example, that ‘Snow is white’ is true iff snow is white. This fact is easy for a deflationary theory to explain because, according to that theory, such beliefs are obvious

once one has mastered the truth term: what is believed is then as obvious as that snow is white iff snow is white (section 6). But how can a correspondence theory explain this striking fact? According to that theory, the beliefs hold in virtue of a certain complicated relation that statements stand in to the world, a relation that the best correspondence theorists have yet to explain satisfactorily. How come so many people innocent of the correspondence theory nonetheless believe instances of the equivalence schema? A deflationist might reasonably claim that there is a strong argument for deflationism here.<sup>32</sup>

I do not have a response to this claim that fully satisfies me. Still, I am dubious about the claim. The correspondence theory holds that the nature of truth is such that the equivalences hold. Certainly, the theory cannot pretend that mastering the truth term teaches you the theory of that nature, any more than it teaches you how to recognize the worldly situations that make statements true. But why can't mastering the truth term teach you the equivalences? For surely mastering the term teaches you that statements are true if the world is as the statements describe. And this, together with some obvious background knowledge, will yield belief in the equivalences. The equivalences do not capture the nature of truth; they are simply a readily apparent manifestation of that nature.<sup>33</sup>

## 10 Explaining Meaning

In light of part I of my discussion, the basic form that a case for correspondence truth must take is clear: truth plays a causal-explanatory role in the world and has a nature that can be explained along correspondence lines. The core of any case for the rival deflationary theory must be a denial of this. For the denial is necessary to motivate the revisionist semantics that the deflationary theory proposes for the truth term. Without the denial, the correspondence theorist can accept that the truth term has the logical role emphasized by the deflationist while insisting that the term has the semantics of a normal descriptive predicate.

Do we need correspondence truth to explain anything? It has been usual to think that correspondence truth has one or more important roles in metaphysics. Correspondence truth is thought to explain individual success, species success, the observational success of scientific theories, and the convergent view of scientific progress. Related to this, correspon-

dence truth is often thought to be essential to realism about the external world. I am skeptical about all this. In particular, I have argued that the issues of realism about the external world and truth have almost nothing to do with each other (1997; but see 2 in the argument below for the interesting exception).

In my view, correspondence truth is important to semantics, not metaphysics. We need correspondence truth to explain *meanings*; more accurately, we need correspondence truth *conditions* to explain meanings. Truth conditions, and hence truth, play a causal role because meanings do. The correspondence truth that is needed is not the traditional one of “correspondence to the facts” but a contemporary one explained in terms of syntactic structures and the references of words that fit into these structures.

I argued for this representational view of meaning in *Coming to Our Senses* (1996a). Here is a summary of the argument:

1. The argument starts by asking, What are meanings (contents) supposed to do? What theoretical purpose do we serve by ascribing them? Meanings are supposed to play at least two very important roles. In virtue of having meanings, beliefs contribute to the explanation of behavior and guide us to the reality that the beliefs concern. Because statements express beliefs (language expresses thought) the meanings of statements have, derivatively, the same two roles.
2. Next consider what folk and social scientists, rightly or wrongly, ascribe to beliefs and statements for these purposes of explaining behavior and being guided to reality; consider what these people, in effect, *take to be* meanings. The properties they ascribe, using ordinary attitude ascriptions like ‘x believes that ...’ and ‘y says that ...’, are, as a matter of fact, entirely constituted by properties that go into determining truth conditions and reference. This is as true when the ascriptions are opaque as when they are transparent. So no “stereotype” or non-reference-determining functional role is ever ascribed. Hence the semantic status quo is truth-referential, and it is not holistic. Furthermore, these truth-referential properties cannot be explained in epistemic terms, on pain of slipping into antirealism (section 1, item (2)). So the properties involve correspondence truth, not verificationist truth.
3. But is the status quo right? We have good reason to think so because these ordinary ascriptions, by folk and social scientists, are mostly successful. So probably the properties they ascribe really do explain behav-

ior and guide us to reality. And so probably those properties really are meanings.

4. Finally, there is no convincing argument for revising this status quo, for thinking that some properties other than the nonepistemic truth-referential ones we ordinarily ascribe would better explain behavior and guide us to reality. There is no good reason to believe in two-factor theories, conceptual-role theories, use theories, and the like. Indeed, these are not so much theories as hand wavings.

At the time I presented this argument, I had not read Brandom's *Making It Explicit* (1994). Brandom's theory is as far from hand waving as one could get. Brandom has always accepted the onus on the deflationist to provide an alternative to the usual truth-referential approach to meaning.<sup>34</sup> In this massive book, he offers a very detailed use theory, drawing on ideas from Kant, Wittgenstein, Sellars, Dummett, and others.<sup>35</sup> The theory gives a "broadly inferential" account, explaining meanings (contents) holistically in terms of inferences, language entries in perception, and language exits in actions. These various "conceptual role" processes are in turn explained in terms of the social practices of undertaking and attributing commitments to think and act in appropriate ways. The practice of attributing commitments, interpreting, is more basic. These social practices are implicitly *normative* in that they implicitly acknowledge the correctness of certain performances. Finally, this implicit normativity is explained in terms of sanctions. However, that does not naturalize the normativity, because the sanctions themselves are "internal" to the normative system.

Brandom is totally up-front about abandoning naturalism. In the preface he notes that his theory "makes essential use of *normative* vocabulary" in specifying use. "No attempt is made to eliminate [this vocabulary], in favor of nonnormative or naturalistic vocabulary" (1994, xiii). Later he describes his order of explanation as the reverse of the traditional naturalistic one, which treats representation as fundamental and hopes to explain "the normative character of the practice in which intentional states are significant ... in ultimately naturalistic terms" (1994, 149).<sup>36</sup>

My own representationalism in *Coming to Our Senses* (1996a) is naturalistic in both the epistemological sense (rejecting a priori knowledge)

and the metaphysical sense (physicalism). And my criticisms of holism, conceptual-role theories, and so on, presuppose naturalism. Perhaps criticisms of this sort could be shown to bear against Brandom without this presupposition. In particular, perhaps he is susceptible to the charge that we in fact do not ever ascribe holistic conceptual-role meanings to explain behavior and guide us to the world. Yet I suspect that from his “interpretative” perspective, he has answers to such criticisms: his book conveys the sense that he has “thought of everything.” So it may be that the objection to Brandom’s theory of meaning is simply that it is not naturalistic. Still, this is a very serious objection because naturalism is worth dying for.

If we hold to naturalism, have we then decided the case for correspondence truth against deflationism? Sadly, no. That case depends not only on correspondence truth’s having an explanatory role in the theory of meaning but also on its having a satisfactory explanation itself. My claim was that it can be explained in terms of syntax and reference. This explanation in turn requires naturalistic explanations of syntax and reference. The latter, at least, has proved very hard to come by. I think that we should be optimistic that some combination of ideas from historical-causal, indicator, and teleological theories will do the trick (Devitt and Sterelny 1999, 156–162). But suppose that this optimism is misplaced. What then? Something appalling: we would have to do what Quine has already done: abandon meaning altogether. But we would not have to abandon truth: we could accept the deflationary theory.

In sum, from a naturalistic perspective, the case for correspondence truth over deflationism is strong if we can explain reference. If we cannot explain reference, then we should adopt deflationism. The very heavy price of this would be eliminating meaning. This price seems so heavy that surely we should be optimistic about explaining reference.

## Notes

Special thanks to Hartry Field, whose doubts about the main theses of this paper have led to many changes. I am grateful for comments at the University of Sydney and the Graduate Center of City University of New York when versions of the paper were presented. My thanks also to Marian David, Paul Horwich, Mark Lance, Bill Lycan, Paul Pietroski and Georges Rey for helpful comments. The



paper is an expanded version of Devitt 2000 and is printed here with the permission of Richard Schantz.

1. For convenience I shall mostly just talk of statements. By 'statement' and 'belief' I refer to meaningful (contentful) *tokens*. Some prefer to talk of the truth of propositions. Nothing I say hinges on which preference is right, so far as I can see.

2. Popper (1968, 274) credits Tarski with making it respectable to engage in truth talk.

3. Realism about truth is not to be confused with realism about the external world (Devitt 1997). Still, realism about truth is like the other realism in having two dimensions. The first of these is an existence dimension, committed, very roughly, to the reality of truth. That is what eliminativists, including deflationists, deny. The second dimension is an independence dimension, committed to the reality of truth being appropriately independent of our minds. The verificationist theory of truth is antirealist in virtue of denying the independence dimension. The correspondence theory, as usually understood and as I am understanding it here, is committed to both dimensions, and so is realist. Still, there could be versions of the correspondence theory without the independence dimension, and according to Richard Kirkham, there have been (1992, 133–134).

4. The "semantics" of a term concerns its meaning. So also does its "linguistics," but the latter may also concern other aspects of the nature and role of the term.

5. See, for example, Devitt 1991; 1997, 3.4.

6. The expressive power that we get from the truth term could also be obtained by introducing sentence variables into our language. See Horwich 1990, 4–5 n., for an interesting discussion.

7. This is rough; for example, we need to guard against the semantic paradoxes and allow for indexicals. The formulation talks of translation because an appropriate instance might refer to a statement that is not in the language of the instance, e.g., "'Schnee ist weiss' is true iff snow is white'.

8. As Mark Lance in effect points out (1997, 184–185).

9. Boghossian (1990a and 1990b) argues that deflationism is inconsistent with nonfactualism about an area and that deflationism itself is incoherent. For some responses, see Devitt 1990, Devitt and Rey 1991, and Soames 1997.

10. Devitt 1996b; 1997, 307–320. The former includes a brief and, it now seems to me, somewhat mistaken discussion of deflationary truth (1996b, 169–170).

11. See, e.g., Ayer 1952, 103, 107; Sayre-McCord 1988, 4–5, 8; Boghossian 1990a, 160–161, 164; Blackburn 1993a, 3, 60; 1993b, 365; Haldane and Wright 1993b, 11–12; Hale 1993, 337, 340.

12. Note that the nonfactualist is not speaking a different language from the factualist. Rather, she has a different theory of the language that they both speak.

13. Concerning realism issues in general, I have argued that, from a naturalistic perspective, we should *always* "put metaphysics first" by establishing a meta-

physical base with near enough no appeal to semantics and by arguing from that base for a semantics. For we know far more about the world than we do about meanings (Devitt 1997; Devitt and Sterelny 1999, 11.4, 12.4).

14. Or that there are any *facts* in the area. This characterization has similar problems to the one about properties. For examples of these characterizations, see Ayer 1952, 89; Wright 1988, 29–30; Sayre-McCord 1988, ix–x, 4; Brandom 1988, 90–91; Boghossian 1990a, 157–159, 161–162; Grover 1992, 14; Blackburn 1993a, 3, 52, 57; Hale 1993, 337; Railton 1993, 280; Soames 1997, 4; Lynch 1998, 112.

15. Kirkham notes the problem that the realism issue about properties poses for the popular characterization. His solution is to characterize deflationism as the thesis “that ‘true’ is not a genuine predicate” (1992, 311). One objection to this characterization is that some deflationists, for example, Horwich, think that ‘true’ is a genuine predicate. A more serious objection is that this characterization is a linguistic one, and we need a metaphysical one.

16. See also Grover 1992, 14. Horwich, however, does claim an explanatory role for truth (1990, 45). I have argued against this view (Devitt 1991, 278–280).

17. Note that we are asking in virtue of what is ‘Snow is white’ true iff snow is white, not in virtue of what is “‘Snow is white’ is true iff snow is white’ true. Deflationism and the correspondence theory would give similarly different responses to the latter question, but the responses would be more complicated.

18. Compare, “All the anaphoric [prosentential] theory of truth tells us about what it is for ‘Snow is white’ to be true, is that it is for snow to be white” (Lance 1997, 188).

19. This is what I have suggested in previous discussions (e.g., Devitt 1991, 276–277; 1997, 32–33). My mistake arose from a use/mention confusion of the sort discussed in the next section.

20. In general, I take it that our *concept* or *notion* of truth can near enough be identified with the meaning of the truth term. In “The Semantic Conception Conception of Truth,” after a similar variety of descriptions of his enterprise, Tarski has this to say about his usage: “The words ‘notion’ and ‘concept’ are used in this paper with all of the vagueness and ambiguity with which they occur in philosophical literature. Thus, sometimes they refer simply to a term, sometimes to what is meant by a term, and in other cases to what is denoted by a term” (1949, 80 n.).

21. A metaphysical definition is exemplified by scientifically defining water as  $H_2O$ ; compare the linguistic definition of ‘vixen’ as ‘female fox’.

22. If a predicate is covered by a description theory, as our example takes ‘bachelor’ to be, it will have a definition. If it is covered by a causal theory, as words like ‘tiger’ very likely are, it will not have a definition. So an explanation of the nature of *F*s can be derived from a theory of ‘*F*’ only if ‘*F*’ is *both* a normal descriptive predicate and covered by a description theory. Even where the explanation of nature can be derived in this way, it should not be, in my view. We

should start with metaphysics, not semantics, because we know more about the world than about meanings (see note 13).

23. The definition of 'designate' does yield 'for  $N$  to designate  $x$  is for  $N$  to be "France" and  $x$  to be France or  $N$  to be "Germany" and  $x$  to be Germany or ...'. Perhaps we could count this as an explanation of the nature of reference: talk of "nature" is not clear enough to rule this out (see section 6). But we cannot count it as a substantial reductive explanation of the sort indicated in section 5.

24. They are also implausible. Brandom (1984) has a more appealing deflationary theory.

25. For example, Rudolf Carnap talks of Tarski's "definition of truth" (1963, 60); Kirkham, in his impressively thorough introduction to theories of truth, nicely distinguishes the metaphysical project from the linguistic one (1992, 20–21) but then places Tarski "firmly with the metaphysical project" (1992, 33).

26. It tells us nothing as it stands, but if we revised it by dropping its listlike definitions, then we could see it as yielding an explanation of truth in terms of reference, as Field points out. If this were then supplemented by a substantial theory of reference, we would have a correspondence theory of truth.

27. Consider also, "The basic idea for deflationary theories of truth ... is roughly that *there is no more to truth than the equivalence thesis*" (Devitt 1997, 30); "The deflationist tells us ...: Truth's 'nature', such as it is, is (pretty much) exhausted by the equivalence of a claim  $p$  with the claim  $p$  is true" (Richard 1997, 57).

28. Horwich is happy to go along with the unselective realist about properties, holding that 'true' is a predicate referring to a "logical" property (1990, 38). So in this respect, instances of the schema are "about truth." But it is still a mistake to think that the instances say anything substantial about the nature of truth. Truth as a logical property has no nature open to reductive explanation. Indeed, it has no properties except trivial ones like being logical and being a property. And although we might perhaps take the equivalence schema to yield an explanation of truth, it does not yield a substantial reductive one (see note 23).

29. On this, see Horwich 1990, 34–37. One must abstract from the conflation of meaning with a speaker's knowledge of meaning (see Devitt and Sterelny 1999, chap. 8).

30. Kirkham takes Horwich at his word and so sees his remarks about the equivalence schema as an answer to the metaphysical question about the nature of truth (Kirkham 1992, 339). As a result of this, and Horwich's acceptance that truth is a property, Kirkham does not classify Horwich as a deflationist.

Soames takes "the leading idea" of deflationism to be that the equivalence schema "is in some sense definitional of the notion of truth" (1997, 4). The talk of "notion" makes this appropriately linguistic. But the talk immediately follows the inappropriately metaphysical: "The equivalences ... are crucial in explaining what truth consists in" (1997, 3). And it is immediately followed by the claim that the statement "There is no such property as truth," which is straightforwardly metaphysical, is a variation of it (1997, 4).

31. Brandom's excellent paper is sadly neglected; it gets no mention, for example, in Kirkham's encyclopedic discussion (1992).
32. For example, see Horwich 1997, 95–96.
33. See also Field 1986, 71–75; Richard 1997, 69–73; Lance 1997, 185–187.
34. In general, Brandom views the truth issue very much as I do, even though we end up with opposite conclusions. Indeed, he might well endorse part I of this paper.
35. Brandom 1997 is a helpful summary.
36. I have offered a brief example of such an explanation of norms in discussing the skeptical position that Kripke finds in Wittgenstein (Devitt and Sterelny 1999, 213–214).

## References

- Armstrong, D. M. 1978. *A Theory of Universals*. Vol. 2, *Universals and Scientific Realism*. New York: Cambridge University Press.
- Ayer, A. J. 1952. *Language, Truth, and Logic*. New York: Dover Publications.
- Blackburn, Simon. 1984. *Spreading the Word*. Oxford: Clarendon Press.
- Blackburn, Simon. 1993a. *Essays in Quasi-Realism*. New York: Oxford University Press.
- Blackburn, Simon. 1993b. "Realism, Quasi, or Queasy?" In Haldane and Wright 1993a, pp. 365–383.
- Boghossian, Paul A. 1990a. "The Status of Content." *Philosophical Review* 99: 157–184.
- Boghossian, Paul A. 1990b. "The Status of Content, Revisited." *Pacific Philosophical Quarterly* 71: 264–278.
- Brandom, Robert. 1984. "Reference Explained Away." *Journal of Philosophy* 81: 469–492.
- Brandom, Robert. 1988. "Pragmatism, Phenomenalism, and Truth Talk." In *Realism and Antirealism*, Midwest Studies in Philosophy, no. 12, ed. Peter A. French, Theodore E. Uehling Jr., and Howard K. Wettstein. Minneapolis: University of Minnesota Press: 75–94.
- Brandom, Robert. 1994. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge: Harvard University Press.
- Brandom, Robert. 1997. "From Truth to Semantics: A Path through *Making It Explicit*." In Villanueva 1997, pp. 141–154.
- Carnap, Rudolf. 1963. "Intellectual Autobiography." In *The Philosophy of Rudolf Carnap*, ed. P. A. Schilpp, pp. 1–84. La Salle: Open Court.
- David, Marian. 1994. *Correspondence and Disquotation: An Essay on the Nature of Truth*. Oxford: Oxford University Press.

- Devitt, Michael. 1990. "Transcendentalism about Content." *Pacific Philosophical Quarterly* 71: 247–263.
- Devitt, Michael. 1991. "Minimal Truth: A Critical Notice of Paul Horwich's *Truth*." *Mind and Language* 6: 273–283.
- Devitt, Michael. 1996a. *Coming to Our Senses: A Naturalistic Program for Semantic Localism*. New York: Cambridge University Press.
- Devitt, Michael. 1996b. "The Metaphysics of Nonfactualism." In *Metaphysics, Philosophical Perspectives*, no. 10, ed. James Tomberlin, pp. 159–176. Atascadero, Calif.: Ridgeview Publishing Co.
- Devitt, Michael. 1997. *Realism and Truth*. 2nd ed. with a new Afterword. Princeton: Princeton University Press. 1st ed., 1984; 2nd ed., 1991.
- Devitt, Michael. 2000. "The Metaphysics of Deflationary Truth." In *What Is Truth?* ed. Richard Schantz. Berlin and New York: Walter de Gruyter.
- Devitt, Michael, and Georges Rey. 1991. "Transcending Transcendentalism: A Response to Boghossian." *Pacific Philosophical Quarterly* 72: 87–100.
- Devitt, Michael, and Kim Sterelny. 1999. *Language and Reality: An Introduction to the Philosophy of Language*. 2nd ed. Oxford: Blackwell Publishers. 1st ed., 1987.
- Field, Hartry. 1972. "Tarski's Theory of Truth." *Journal of Philosophy* 69: 347–375.
- Field, Hartry. 1986. "The Deflationary Conception of Truth." In *Fact, Science, and Morality: Essays on A. J. Ayer's Language, Truth, and Logic*, ed. Graham Macdonald and Crispin Wright, pp. 55–117. Oxford: Basil Blackwell.
- Grover, Dorothy. 1992. *A Prosentential Theory of Truth*. Princeton: Princeton University Press.
- Haldane, John, and Crispin Wright, eds. 1993a. *Reality, Representation, and Projection*. New York: Oxford University Press.
- Haldane, John, and Crispin Wright. 1993b. "Introduction." In Haldane and Wright 1993a, pp. 3–12.
- Hale, Bob. 1993. "Can There Be a Logic of Attitudes?" In Haldane and Wright 1993a, pp. 337–363.
- Horwich, Paul. 1990. *Truth*. Oxford: Basil Blackwell.
- Horwich, Paul. 1997. "Deflationary Theory and the Problem of Aboutness." In Villanueva 1997, pp. 95–106.
- Kirkham, Richard. 1992. *Theories of Truth: A Critical Introduction*. Cambridge: MIT Press.
- Lance, Mark. 1997. "The Significance of Anaphoric Theories of Truth and Reference." In Villanueva 1997, pp. 181–198.
- Leeds, Stephen. 1978. "Theories of Reference and Truth." *Erkenntnis* 13: 111–129.

- Lynch, Michael P. 1998. *Truth in Context: An Essay on Pluralism and Objectivity*. Cambridge: MIT Press.
- Popper, Karl. 1968. *Logic of Scientific Discovery*. New York: Basic Books.
- Quine, W. V. 1970. *Philosophy of Logic*. Englewood Cliffs: Prentice-Hall.
- Railton, Peter. 1993. "What the Non-cognitivist Helps Us to See the Naturalist Must Help Us to Explain." In Haldane and Wright 1993, pp. 279–300.
- Richard, Mark. 1997. "Deflating Truth." In Villanueva 1997, pp. 57–78.
- Sayre-McCord, Geoffrey. 1988. Preface and "Introduction: The Many Moral Realisms." In *Essays on Moral Realism*, ed. Geoffrey Sayre-McCord, pp. ix–xii, 1–23. Ithaca: Cornell University Press.
- Soames, Scott. 1984. "What Is a Theory of Truth?" *Journal of Philosophy* 81: 411–429.
- Soames, Scott. 1997. "The Truth about Deflationism." In Villanueva 1997, pp. 1–44.
- Tarski, Alfred. 1949. "The Semantic Conception of Truth and the Foundations of Semantics." In *Readings in Philosophical Analysis*, ed. H. Feigl and W. Sellars, pp. 52–84. New York: Appleton-Century-Crofts.
- Tarski, Alfred. 1956. *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*. Oxford: Clarendon Press.
- Villanueva, Enrique, ed. 1997. *Truth*. Philosophical Issues, no. 8. Atascadero: Ridgeview Publishing Co.
- Wright, Crispin. 1988. "Realism, Antirealism, Irrealism, Quasi-Realism." In *Realism and Antirealism*, Midwest Studies in Philosophy, no. 12, ed. Peter A. French, Theodore E. Uehling Jr., and Howard K. Wettstein, pp. 25–49. Minneapolis: University of Minnesota Press.



## VII

---

### Primitivism, Identity Theory, and Alethic Pluralism





---

# Primitivism, Identity Theory, and Alethic Pluralism: Introduction

Even a fairly quick perusal of most of the essays in this volume can leave one with the impression that the debate over truth has reached an impasse. On one side are the robust accounts of truth, such as the correspondence and coherence theories. On the other is deflationism and its varieties. Since both approaches are riddled with problems, one senses a stalemate.

In one way or another, all of the essays in this section discuss ways to think about truth that look beyond the traditional theories. A common thread is that the failure of substantive definitions of truth needn't lead to a thoroughgoing deflationism. Broadly speaking, these authors discuss three alternatives: primitivism, or taking truth as a basic, indefinable concept; the identity theory, which identifies true propositions with facts; and pluralism, which takes truth to have different natures in different discourses.

## Primitivism

G. E. Moore famously held that *good* was a simple unanalyzable concept. Like the concept of yellow, Moore argued that goodness could not be defined by any simpler concepts. The concept of good is in this sense basic. It is less well known that Moore (and Russell) briefly flirted with a similar view about truth, namely that truth “is a simple unanalyzable property which is possessed by some propositions and not by others” (Moore 1953).

Although later abandoned by Moore, the view that truth is a basic or primitive concept has recently seen something of a comeback. Its principle advocates are Donald Davidson and Ernest Sosa. In both philosophers’

views, previous attempts to supply substantive content to the concept of truth—attempts like those of the correspondence and coherence theories, for example—are either subject to counterexample or devoid of content themselves. This is exactly what primitivism tells us to expect; these theories have failed to define truth because it can't be defined. Davidson adds two other points. First, he notes that our inability to reduce truth to a more basic set of concepts is hardly surprising. The concept of truth is already so basic to our thought that without it we might not have any concepts at all. Second, Davidson takes Tarski's work (chap. 15) to prove that the truth predicate is undefinable. Tarski shows that we can apply the truth predicate to the sentences of a language only if we relativize its application to that language. Thus Tarski defines not "true" but "true-in-L," "true-in-L<sub>2</sub>," and so on. About truth itself, Davidson argues, Tarski can say nothing.

Both Davidson and Sosa are concerned to distinguish primitivism from deflationism, and in particular, Horwich's minimalist theory (chap. 24). The minimalist also believes in the failure of the traditional theories, and also takes the truth predicate to be immune from explicit definition. But primitivism is consistent with taking truth to be an explanatorily important concept, while minimalism is not. Sosa forcefully argues that the concept of truth is needed to understand the central concepts of epistemology, such as justification and the notion of an intellectual virtue. Davidson, on the other hand, takes truth to be crucial to meaning; without an idea of truth, we cannot understand meaning in terms of truth conditions.

While both Sosa and Davidson believe that the truth predicate can't be defined, they disagree about what more can be said about truth. For Sosa, primitivism about the concept of truth is consistent with a substantive metaphysical theory that takes truth as involving correspondence with reality—so long as that theory is not seen as implying anything about the concept of truth. Davidson, on the other hand, eschews any sort of robust metaphysical explanation of truth. For Davidson, what else we can say about truth is how that concept relates to other concepts, our attitudes, and our behavior. In this sense, Davidson sees the project of giving what he calls "empirical content" to truth as inseparable from the project of assigning content to any of our observable linguistic behavior.

## The Identity Theory

Jennifer Hornsby also takes it that truth cannot be analyzed in any traditional sense. Nonetheless, she does think there is something metaphysically robust to say about truth. Drawing on some suggestions from John McDowell, she argues that true thinkables are the same as facts. This could be called an identity theory of truth.

By “thinkables,” Hornsby means the contents of our thought, as opposed to the thinking. If one takes the content of my belief that *p* to be a proposition, then the identity theory amounts to the view that true propositions are identical with facts. This means that the theory is not an epistemic theory; it is not required that true propositions be justifiedly believed or known. But by the same token, truth does not—as it does on the correspondence theory—involve a relation between a proposition and a truth maker of some sort. There is no gap between true thinkables and the way the world is; indeed, the mind and the world seem to envelop each other on the identity account. And yet the theory is committed to the existence of the property of truth, and hence it is not a deflationary view in the ordinary sense of the word either.

Marian David discusses two lines of objection to the identity theory. One is the problem of falsehood. If *x* is a true proposition if and only if *x* is a fact, then how do we understand what it means for *x* to be a false proposition? As David points out, we cannot say “when *x* is not a fact,” because there are presumably some things, such as my computer, which are neither facts nor propositions. Thus David suggests that the identity theorist must allow for some propositions that, while being entirely similar to facts in structure, are not facts. All facts will be (true) propositions, but not all propositions will be facts. But this solution raises the other problem that David discusses, namely the question of how the identity theorist understands the nature of propositions or content. For clearly, how the identity theorist sees propositions will determine how she sees facts as well. Finally, David argues that the identity theory does not necessarily oppose the correspondence theory. It can be seen as a component of a correspondence theory of sentential truth. On such a view, a sentence is true when it represents a true proposition, and true propositions are identical to facts.

## Pluralist Theories of Truth

In the selection included here from his 1994 Dewey Lectures, Hilary Putnam argues against his previous endorsement of Dummett's verificationist semantics (see chaps. 10 and 11). According to Putnam, Dummett's verificationism results in a "loss of the world," in particular, by implying that there can be no facts of the matter concerning the distant past.

Yet Putnam also argues against the usual alternatives to Dummett's antirealist picture: deflationism and metaphysical realism. Putnam claims that far from being a real alternative to antirealism, deflationism must actually presuppose a verificationist account of understanding. The deflationist rejects the idea that truth is a substantive explanatory property (see the introduction to part VI). As such, they can't explain the meaning of sentences in terms of their truth conditions, for to do so is to assume an explanatory role for the property of truth. Instead, they must take meaning to consist in the conditions under which a sentence would be justifiably asserted—when it would be verified, in other words. So the deflationist ends up presupposing verificationism.

The metaphysical realist is right to reject verificationism, Putnam believes, but wrong in thinking that there must be a single substantive property that every true proposition shares no matter what the context. Instead, Putnam argues, there is a plurality of ways for propositions to relate to reality. The word "true," he suggests, has different uses, depending on whether we are talking about morality, mathematics, physical reality, and so on.

Putnam's pluralist conception of thought and truth is intriguing. But we might wonder what it amounts to at the detailed level. One question, for instance, is this: does Putnam, in saying that "true" has a variety of uses, mean to imply that the word is systematically ambiguous? If so, how can we explain the fact that we can make generalizations about truth, such as the generalization that every proposition of the form "If  $p$ , then  $p$ " is true?

Crispin Wright presents a case for pluralism about truth by arguing for what he calls a minimalist view of the concept. His position is minimalist

in the sense that, like deflationism, his view takes it that our concept of truth is completely captured by its adherence to principles like the equivalence schema (“It is true that  $p$  if and only if  $p$ ”). But it is important to see that Wright’s version of minimalism is distinct from the deflationary view of the same name defended by Paul Horwich (chap. 24). For while Wright is a minimalist about the *concept* of truth, he takes truth as a robust *property* of propositions. Indeed, in Wright’s view, deflationism ultimately fails because it cannot account for the fact that truth acts as a substantive norm—distinct from warranted assertibility—for our thought and talk.

Like many of the other authors above, Wright eschews a traditional analysis of the concept of truth in terms of necessary and sufficient conditions. Instead, he suggests what he calls an “analytical theory” of the concept. Such a theory consists in an assemblage of platitudes about truth: basic principles that we take to hold a priori of any true proposition. This family of principles serves to mark out a single concept of truth. But this way of construing the concept leaves room for a plurality of truth properties. For, Wright argues, these principles—and therefore the concept of truth—can be realized or satisfied by different underlying properties in different discourses.

Much of Wright’s rich discussion is devoted to the question of how best to capture the spirit of the pragmatist conception of truth (see Part III). After discussing and ultimately rejecting Putnam’s view on the matter (chap. 11), Wright suggests that the basic insight behind pragmatism can be captured by the idea of *superassertibility*. Roughly speaking, a proposition is superassertible just when it is enduringly justified. That is, a superassertible proposition is one whose justification is never defeated or overridden by future evidence. Wright argues that superassertibility can satisfy the various platitudes that constitute our analytical theory of truth in certain domains of discourse. In particular, Wright claims, superassertibility may constitute truth in any discourse where it is a priori that all the truths of that discourse are knowable. Wright suggests moral and comic discourse as possible examples. The implication of this position is clear. *Some* truth, in Wright’s view, needn’t be understood in terms of a relation between propositions and objects; e.g., moral truth can be understood as an epistemic property of moral propositions.

It is worth noting in *this* context that in chap. 4 Terence Horgan presents a version of the correspondence theory of truth with strong affinities to both Wright's view and Putnam's present positions. According to Horgan, truth in any discourse is what he calls "semantic correctness." To say that *p* is semantically correct is *not* to say that it is justifiable—in Horgan's view, semantic norms are not reducible to epistemic norms. Nonetheless, the semantic norms that govern correct assertibility (i.e., truth) *can* vary from one discourse to another. On this picture, the sentences of some discourses will directly correspond to reality, while those of other discourses will only indirectly correspond to the world. Yet Horgan stops short, it seems, of embracing a fully pluralist account of truth (see the introduction to part I).

In "A Functionalist Theory of Truth" I express sympathy with the pluralist theories just discussed, and with Wright's theory in particular. But I argue that the cause of pluralism is best served by explicitly construing truth as a functional concept. In this view, alethic concepts like truth, fact, and reference are similar to psychological concepts as those concepts are understood by the (analytic) functionalist theory of the mind. The psychological functionalist takes the concept of pain to be the concept of a functional, or multiply realizable, property of organisms. To be in pain is to have a lower-order neural property that plays the role of pain; thus "being in pain" means that one has the higher-order property of having a property that plays the pain role. A functionalist theory of truth, therefore, takes the concept of truth in every context as the concept of a higher-order functional property of propositions. But that property, by its very nature, can be realized differently in different contexts. Thus propositions about the physical world might be true in virtue of their correspondence with mind-independent objects, while the truth of what I call purely legal propositions might be realized by superassertibility. The functional role of truth is defined in terms of the place it occupies in a certain network of principles and platitudes. Taken together, the principles of this network can be used to give a rigorous job description for every alethic concept all at once.

Pluralist theories of truth have significant advantages. Most important, they account for the fact that every traditional theory of truth seems plausible in some domains but not in others. But they also raise serious

concerns. For one thing, they seem to imply a pluralism not just about truth but about every philosophical concept related to truth. Arguably, knowledge, meaning, and logic will all be affected. Further, some discourses, such as moral and legal discourse, are deeply intertwined with each other. Many propositions may not fit clearly into any discourse. Thus pluralist theories, no matter what their details, must give us a way to understand what makes one type of “discourse” differ from another.

### Further Reading for Part VII

#### Primitivism

Cartwright, R. 1987. “A Neglected Theory of Truth.” In his *Philosophical Essays*, pp. 71–95. Cambridge: MIT Press.

Davidson, D. 1990. “The Structure and Content of Truth.” *Journal of Philosophy* 87: 279–238.

Moore, G. E. 1953. *Some Main Problems in Philosophy*. London: Allen and Unwin.

#### The Identity Theory

Baldwin, T. 1991. “The Identity Theory of Truth.” *Mind* 100: 35–52.

Candlish, S. 1995. “Resurrecting the Identity Theory of Truth.” *Bradley Studies* 1: 116–124.

Dodd, J. 1995. “McDowell and Identity Theories of Truth.” *Analysis* 55: 160–165.

McDowell, J. 1994. *Mind and World*. Cambridge: Harvard University Press.

Moore, G. E. 1899. “The Nature of Judgement.” *Mind* 8: 176–193.

#### Pluralism about Truth

Blackburn, S. 1998a. “Wittgenstein, Wright, Rorty, and Minimalism.” *Mind* 107: 158–182.

Lynch, M. 1998. *Truth in Context*. Cambridge: MIT Press.

Pettit, Philip. 1996. “Realism and Truth: A Comment on Crispin Wright’s Truth and Objectivity.” *Philosophy and Phenomenological Research* 56, no. 4: 883–889.

Sainsbury, M. 1996. “Crispin Wright: Truth and Objectivity.” *Philosophy and Phenomenological Research* 56, no. 4: 899–904.

Wright, C. 1995. “Truth in Ethics.” *Ratio* (new series) 8, no. 3: 210–226.

Wright, C. 1996. “Response to Commentators.” *Philosophy and Phenomenological Research* 56, no. 4: 911–941.

Wright, C. 1998. “Comrades against Quietism.” *Mind* 107: 184–203.





---

## The Folly of Trying to Define Truth

Donald Davidson

In the *Euthyphro*, Socrates asks what holiness is, what “makes” holy things holy. It is clear that he seeks a definition, a definition with special properties. He spurns the mere provision of examples or lists, asking in each case what makes the examples examples, or puts an item on the list. He rejects merely coextensive concepts (“something is holy if and only if it is dear to the gods”): what makes something dear to the gods is that it is holy, but not vice versa. The dialogue ends when Socrates begs Euthyphro to enlighten him by coming up with a satisfactory answer; Euthyphro decides he has another appointment.

The pattern of attempted definition, counterexample, amended definition, further counterexample, ending with a whimper of failure, is repeated with variations throughout the Socratic and middle Platonic dialogues. Beauty, courage, virtue, friendship, love, temperance are put under the microscope, but no convincing definitions emerge. The only definitions Plato seems happy with are tendentious characterizations of what it is to be a sophist. He also gives a few trivial samples of correct definitions: of a triangle; of mud (earth and water).

In the *Theaetetus*, Plato attempts to define empirical knowledge. Like many philosophers since, he takes knowledge to be true belief plus something more—an account that justifies or warrants the belief. It is the last feature which stumps him (again foreshadowing the subsequent history of the subject). It seems no more to occur to Plato than it has to most others that the combination of causal and rational elements that must enter into an analysis of justified belief (as it must into accounts of memory, perception, and intentional action) may in the nature of the case not be amenable to sharp formulation in a clearer, more basic, vocabulary.

What is important in the present context, however, is the fact that in attempting to define knowledge, it is only with the concept of warrant that Plato concedes defeat. He does not worry much about the equal involvement of knowledge with truth and belief.

Again, though, Plato was simply blazing a trail that other philosophers over the ages have followed: you follow his lead if you worry about the concept of truth when it is the focus of your attention, but you pretend you understand it when trying to cope with knowledge (or belief, memory, perception, and the like). We come across the same puzzling strategy in David Hume and others, who forget their skepticism about the external world when they formulate their doubts concerning knowledge of other minds. When a philosopher is troubled by the idea of an intentional action, he would be happy if he could analyze it correctly in terms of the concepts of belief, desire, and causality, and he does not for the moment worry too much about those (at least equally difficult) concepts. If memory is up for analysis, the connections with belief, truth, causality, and perhaps perception, constitute the problem, but these further concepts are *pro tem* taken to be clear enough to be used to clarify memory, if only the connections could be got right. It is all right to assume you have an adequate handle on intention and convention if your target is meaning. I could easily go on.

There is a lesson to be learned from these familiar, though odd, shifts in the focus of philosophical puzzlement. The lesson I take to heart is this: however feeble or faulty our attempts to relate these various basic concepts to each other, these attempts fare better, and teach us more, than our efforts to produce correct and revealing definitions of basic concepts in terms of clearer or even more fundamental concepts.

This is, after all, what we should expect. For the most part, the concepts philosophers single out for attention, like truth, knowledge, belief, action, cause, the good and the right, are the most elementary concepts we have, concepts without which (I am inclined to say) we would have no concepts at all. Why then should we expect to be able to reduce these concepts definitionally to other concepts that are simpler, clearer, and more basic? We should accept the fact that what makes these concepts so important must also foreclose on the possibility of finding a foundation for them which reaches deeper into bedrock.

We should apply this obvious observation to the concept of truth: we cannot hope to underpin it with something more transparent or easier to grasp. Truth is, as G. E. Moore, Bertrand Russell, and Gottlob Frege maintained, and Alfred Tarski proved, an indefinable concept. This does not mean we can say nothing revealing about it: we can, by relating it to other concepts like belief, desire, cause, and action. Nor does the indefinability of truth imply that the concept is mysterious, ambiguous, or untrustworthy.

Even if we are persuaded that the concept of truth cannot be defined, the intuition or hope remains that we can characterize truth using some fairly simple formula. What distinguishes much of the contemporary philosophical discussion of truth is that though there are many such formulas on the market, none of them seems to keep clear of fairly obvious counterexamples. One result has been the increasing popularity of minimalist or deflationary theories of truth—theories that hold that truth is a relatively trivial concept with no “important connections with other concepts such as meaning and reality.”

I sympathize with the deflationists; the attempts to pump more content into the concept of truth are not, for the most part, appealing. But I think the deflationists are wrong in their conclusion, even if mostly right in what they reject. I shall not pause here to give my reasons for refusing to accept correspondence theories, coherence theories, pragmatic theories, theories that limit truth to what could be ascertained under ideal conditions or justifiably asserted, and so on.<sup>2</sup> But since I am with the deflationists in being dissatisfied with all such characterizations of truth, I shall say why deflationism seems to me equally unacceptable.

Aristotle, as we all know, contended that

- (1) To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, or of what is not that it is not, is true.

When Tarski<sup>3</sup> mentions this formulation in 1944, he complains that it is “not sufficiently precise and clear,” though he prefers it to two others:

- (2) The truth of a sentence consists in its agreement with (or correspondence to) reality.

- (3) A sentence is true if it designates an existing state of affairs (ibid., p. 343).

In 1969, Tarski<sup>4</sup> again quotes (1), and adds,

[T]he formulation leaves much to be desired from the point of view of precision and formal correctness. For one thing, it is not general enough; it refers only to sentences that “say” about something “that it is” or “that it is not”; in most cases it would hardly be possible to cast a sentence in this mold without slanting the sense of the sentence and forcing the spirit of the language (ibid., p. 63).

He adds that this may be the reason for such “modern substitutes” for Aristotle’s formulations as (2) and (3).

In the *Wahrheitsbegriff*, however, Tarski<sup>5</sup> prefers the following informal statement:

- (4) A true sentence is one which says that the state of affairs is so and so, and the state of affairs indeed is so and so (ibid., p. 155).

It seems to me that Aristotle’s formulation is clearly superior to (2), (3), and (4); it is more in accord with Tarski’s own work on truth; and Tarski’s comment that (1) is “not general enough” is strangely out of keeping with the spirit of his own truth definitions.

(1) is superior to (2)–(4) for three reasons. First, (3) and (4) mention states of affairs, thus suggesting that postulating entities to correspond to sentences might be a useful way of characterizing truth. (“A true sentence is one that corresponds to the facts,” or “If a sentence is true, there is a state of affairs to which it corresponds.”) But facts or states of affairs have never been shown to play a useful role in semantics, and one of the strongest arguments for Tarski’s definitions is that in them nothing plays the role of facts or states of affairs. This is not surprising, since there is a persuasive argument, usually traced to Frege (in one form) or Kurt Gödel (in another), to the effect that there can be at most one fact or state of affairs. (This is why Frege said all true sentences name the True.) Tarski’s truth definitions make no use of the idea that a sentence “corresponds” to anything at all. We should not take seriously the mention of “states of affairs” in such remarks of Tarski’s<sup>6</sup> as this: “[S]emantical concepts express certain relations between objects (and states of affairs) referred to in the language discussed and expressions of the language referring to those objects” (ibid., p. 403).

A second reason for preferring Aristotle's characterization of truth is that it avoids the awkward blanks marked by the words 'so and so' in Tarski's version (4); one is hard pressed to see how the blanks are to be filled in. Aristotle's formula, on the other hand, sounds much like a generalization of Tarski's convention-T.

The third reason for preferring Aristotle's characterization is that it makes clear, what the other formulations do not, that the truth of a sentence depends on the inner structure of the sentence, that is, on the semantic features of the parts. In this it is once again closer to Tarski's approach to the concept of truth.

Tarski's convention-T, which he understandably substitutes for the rough formulas I have been discussing, stipulates that a satisfactory definition of a truth predicate 'is true' for a language  $L$  must be such as to entail as theorems all sentences of the form

$s$  is true-in- $L$  if and only if  $p$

where ' $s$ ' is replaced by the description of a sentence, and ' $p$ ' is replaced by that sentence, or a translation of the sentence into the meta-language. Since it is assumed that there is an infinity of sentences in  $L$ , it is obvious that, if the definition of the truth predicate is to be finite (Tarski insisted on this), the definition must take advantage of the fact that sentences, though potentially infinite in number, are constructed from a finite vocabulary. For the languages Tarski considered, and for which he showed how to define truth, all sentences can be put into the form of an existential quantification, or the negation of an existential quantification, or a truth-functional compound of such sentences. So how "incomplete," from Tarski's point of view, is Aristotle's formulation (1)? It deals with four cases. There are the sentences that "say of what is that it is not": in modern terms it is a false sentence that begins 'It is not the case that there exists an  $x$  such that ...'. An example might be: 'There does not exist an  $x$  such that  $x = 4$ '. Then there are sentences that "say of what is not that it is"; for example: 'There exists an  $x$  such that  $x = 4$  &  $x = 5$ '. There are sentences that "say of what is that it is"; for example: 'There exists an  $x$  such that  $x = 4$ '. And, finally, there are sentences that "say of what is not that it is not"; for example, 'It is not the case that there exists an  $x$  such that  $x \neq x$ '. According to the classical formulation,

sentences of the first two kinds are false and of the second two kinds are true. Tarski is so far in agreement. What would Tarski add? Just the truth-functional compounds (beyond those involving negation) of the types of sentences already mentioned; these are true or false on the basis of the truth or falsity of the kinds of sentences already provided for. Of course, Tarski also showed in detail how the truth or falsity of the first four types of sentences depended in turn on their structure.

Thus, the classical formulation regarded as an informal characterization is “incomplete” in only a minimal way compared to Tarski’s own work, and is better than Tarski’s informal attempts to state the intuitive idea. Needless to say, someone might question the extent to which natural languages can be adequately characterized using such limited resources; but this is a comment equally applicable to Tarski.

Despite his nod in the direction of a correspondence theory, in which sentences are said to correspond to facts, Tarski ought not to be considered as giving comfort to serious partisans of correspondence theories, nor should Aristotle. For neither Aristotle’s formula nor Tarski’s truth definitions introduce entities like facts or states of affairs for sentences to correspond to. Tarski does define truth on the basis of the concept of satisfaction, which relates expressions to objects, but the sequences that satisfy sentences are nothing like the “facts” or “states of affairs” of correspondence theorists, since if one of Tarski’s sequences satisfies a closed sentence, thus making it true, then that same sequence also satisfies every other true sentence, and thus also makes it true, and if any sequence satisfies a closed sentence, every sequence does.<sup>7</sup>

If Tarski is not a correspondence theorist (and he certainly does not hold a coherence theory or a pragmatic theory or a theory that bases truth on warranted assertability), is he a deflationist? Here opinions differ widely: W. V. Quine thinks he is, and so does Scott Soames. John Etchemendy thinks Tarski simply says nothing about truth as a semantic concept, and Hilary Putnam, though for somewhat different reasons, agrees.<sup>8</sup>

If Tarski has said “all there is to say” about truth, as Stephen Leeds, Paul Horwich, and Soames all contend, and Quine has strongly hinted, then a sort of deflationary attitude is justified; this is not quite the same as the “redundancy” view, but close to it. The redundancy view, taken

literally, is the same as the disquotational view taken literally: we can always substitute without loss a sentence for that same sentence quoted, and followed by the words 'is true'. What Tarski added, as Michael Williams and others have pointed out, is a way of predicating truth of whole classes of sentences, or of sentences to which we do not know how to refer; you may think of this as an elaboration of the redundancy theory in that it allows the elimination of the truth predicate when applied to sentences of a language for which that predicate has been defined.

At the same time that we credit Tarski with having shown how to make sense of remarks like 'The English sentence Joan uttered about Abbot was true' or 'Everything Aristotle said (in Greek) was false' or 'The usual truth table for the conditional makes any conditional true that has a false antecedent', we have to recognize that this accomplishment was accompanied by a proof that truth cannot (given various plausible assumptions) be defined in general; there can be no definition of 'For all languages  $L$ , and all sentences  $s$  in  $L$ ,  $s$  is true in  $L$  if and only if ...  $s$  ...  $L$  ...'. In other words, Tarski justified the application of a truth predicate to the sentences of a particular language only by restricting its application to the sentences of that language. (It is ironic that in much recent writing on deflationary theories, Tarski has been taken to have lent support to the idea that there is a single, simple, even trivial, concept of truth.)

A deflationary attitude to the concept of truth is not, then, encouraged by reflection on Tarski's work. One can adopt the line advanced by Putnam and Etchemendy that Tarski was not even doing semantics, despite his insistence that he was; but this construal of Tarski does not support a deflationary theory: it simply denies the relevance of Tarski's results to the ordinary concept of truth. If, on the other hand, one takes Tarski's truth definitions to say something about the relations of specific languages to the world, one cannot at the same time claim that he has told us all there is to know about the concept of truth, since he has not told us what the concept is that his truth definitions for particular languages have in common.

I think that Tarski was not trying to define *the* concept of truth—so much is obvious—but that he was *employing* that concept to characterize the semantic structures of specific languages. But Tarski did not indicate



how we can in general reduce the concept of truth to other more basic concepts, nor how to eliminate the English predicate 'is true' from all contexts in which it is intelligibly applied to sentences. Convention-T is not a rough substitute for a general definition: it is part of a successful attempt to persuade us that his formal definitions apply our single pre-theoretical concept of truth to certain languages. Deflationists cannot, then, appeal to Tarski simply because he demonstrated how to handle the semantics of quantification for individual languages. Leeds, Horwich, Williams, and others who have contended that all Tarski did was reveal the usefulness of an otherwise dispensable concept are wrong. They are right that we need a truth predicate for the purposes they, along with Tarski, mention; but they fail to note the obvious fact that at the same time Tarski solved one problem he emphasized another: that he had not, and could not, given the constraints he accepted, define or fully characterize truth.

Over the years, Quine has said a number of things about truth, but there has been, from early days until the most recent, what seems a consistent embrace of a deflationary attitude. Thus, Quine has made much of the "disquotational" aspect of the truth predicate, the fact that we can get rid of the predicate 'is true' after the quotation of an English sentence simply by removing the quotation marks as we erase the truth predicate. As Quine put it in *From a Logical Point of View*,<sup>9</sup> we have a general paradigm, namely,

(T) '\_\_\_\_' is true-in-*L* if and only if \_\_\_\_

which, though not a definition of truth, serves to endow 'true-in-*L*' with every bit as much clarity, in any particular application, as is enjoyed by the particular expressions of *L* to which we apply [it]. Attribution of truth in particular to 'Snow is white' ... is every bit as clear to us as attribution of whiteness to snow (ibid., p. 138).

In *Word and Object*, Quine<sup>10</sup> remarks that "To say that the statement 'Brutus killed Caesar' is true, or that 'The atomic weight of sodium is 23' is true, is in effect simply to say that Brutus killed Caesar, or that the atomic weight of sodium is 23" (ibid., p. 24). The theme is repeated thirty years later in *Pursuit of Truth*:<sup>11</sup>

there is surely no impugning the disquotation account; no disputing that “Snow is white” is true if and only if snow is white. Moreover, it is a full account; it explicates clearly the truth or falsity of every clear sentence. (ibid., p. 93).

“Truth,” he summarizes, “is disquotation” (ibid., p. 80). On this matter, Quine has not changed his mind.

It is the disquotational feature of truth, in Quine’s opinion, which makes truth so much clearer a concept than meaning. Comparing theory of meaning and theory of reference, Quine says that they constitute “two provinces so fundamentally distinct as not to deserve a joint appellation at all.”<sup>12</sup> The former deals with such tainted topics as synonymy, meaning, and analyticity. The concepts treated by the latter, which include truth, are by contrast “very much less foggy and mysterious...” For although ‘true-in-*L*’ for variable ‘*L*’ is not definable, “what we do have suffices to endow ‘true-in-*L*’, even for variable ‘*L*’, with a high enough degree of intelligibility so that we are not likely to be averse to using the idiom” (ibid., pp. 137–38). “What we do have” is, of course, the paradigm (T) and the “expedient general routine” due to Tarski for defining ‘true-in-*L*’ for particular languages.

The disquotational feature of truth, wedded to the thought that this may exhaust the content of the concept of truth, encourages the idea that truth and meaning can be kept quite separate. But can they in general? Scattered remarks in Quine’s work suggest otherwise. In 1936, Quine published the brilliant and prescient “Truth by Convention.”<sup>13</sup> In it he remarks that “in point of meaning ... a word may be said to be determined to whatever extent the truth or falsehood of its contexts is determined” (ibid., p. 89). It is hard to see how truth could have this power of determining meaning if the disquotational account were all there were to say about truth. Other passages in Quine suggest the same idea: “First and last, in learning language, we are learning how to distribute truth values. I am with Davidson here; we are learning truth conditions.”<sup>14</sup> Or again, “Tarski’s theory of truth [is] the very structure of a theory of meaning.”<sup>15</sup>

Up to a point it may seem easy to keep questions of truth and questions of meaning segregated. Truth we may think of as disquotational (in the extended Tarski sense) and therefore trivial; meaning is then another matter, to be taken care of in terms of warranted assertability, function,

or the criteria for translation. This is the line followed, for example, by Horwich in his recent book *Truth* (op. cit.), by Soames,<sup>16</sup> and by Lewis.<sup>17</sup> It may, at least at one time, have been Quine's view. In *Word and Object*, in a passage that immediately precedes the remark that to say that the sentence 'Brutus killed Caesar' is true is in effect simply to say that Brutus killed Caesar, Quine despairs of a substantive concept of truth, and concludes that we make sense of a truth predicate only when we apply it to a sentence "in the terms of a given theory, and seen from within the theory" (op. cit., p. 24). This is, I think, what Quine means when he says that truth is "immanent." The point is not merely that the truth of a sentence is relative to a language; it is that there is no transcendent, single concept to be relativized.<sup>18</sup>

Most recently, however, Quine muses that truth "is felt to harbor something of the sublime. Its pursuit is a noble pursuit, and unending"; he seems to agree: "Science is seen as pursuing and discovering truth rather than as decreeing it. Such is the idiom of realism, and it is integral to the semantics of the predicate 'true'."<sup>19</sup>

I turn now to Horwich's version of deflationism, for he seems to me to have accepted the challenge other deflationists have evaded, that of saying something more about an unrelativized concept of truth than we can learn from Tarski's definitions. Horwich's brave and striking move is to make the primary bearers of truth propositions—not exactly a new idea in itself, but new in the context of a serious attempt to defend deflationism. He is clear that he cannot provide an explicit definition of a truth predicate applying to propositions, but he urges that we really have said all there is to know about such a predicate (and hence the predicate it expresses) when we grasp the fact that the "uncontroversial instances" of the schema:

The proposition that  $p$  is true if and only if  $p$

exhaust its content. (The limitation to "uncontroversial instances" is to exclude whatever leads to paradox.) The schema is taken as an axiom schema: the totality of its instances constitute the axioms of his theory.

This theory is, of course, incomplete until the controversial instances are specified in a non-question-begging way; and since the set of axioms is infinite, it does not meet one of Tarski's requirements for a satisfactory

theory of truth. But perhaps the first difficulty can be overcome, and the second may be viewed as the price of having an unrelativized concept of truth. There are, further, the doubts many of us have about the existence of propositions, or at least of the principles for individuating them.

All these considerations give me pause, but I plan to ignore them here. I want to give deflationism its best chance, since it seems to me to be the only alternative to a more substantive view of truth, and most substantive views are in my opinion, as in Horwich's, clear failures. But although I enthusiastically endorse his arguments against correspondence, coherence, pragmatic, and epistemic theories, I cannot bring myself to accept Horwich's "minimal" theory.

I have two fundamental problems with Horwich's theory, either of which alone is reason to reject it if it cannot be resolved; and I do not myself see how to resolve them.

The first problem is easy to state: I do not understand the basic axiom schema or its instances. It will help me formulate my difficulty to compare Horwich's axiom schema with Tarski's informal (and ultimately supplanted) schema:

'\_\_\_\_' is true if and only if \_\_\_\_

Tarski's objection (among others) is that you cannot turn this into a definition except by quantifying into a position inside quotation marks. The complaint ends up with a question about the clarity of quotations: How does what they refer to depend on the semantic properties of their constituents? It has sometimes been proposed to appeal to substitutional quantification, and one may wonder why Horwich cannot generalize his schema:

(*p*)(the proposition that *p* is true if and only if *p*)

by employing substitutional quantification. But here Horwich quite rightly explains that he cannot appeal to substitutional quantification to explain truth, since substitutional quantification must be explained by appeal to truth.

Why, though, does Horwich not try generalizing his schema by quantifying over propositions? The answer should be: because then we would have to view ordinary sentences as singular terms *referring* to propositions, not as *expressing* propositions. This brings me to the crux: How

are we to understand phrases like ‘the proposition that Socrates is wise’? In giving a standard account of the semantics of the sentence ‘Socrates is wise’, we make use of what the name ‘Socrates’ names, and of the entities of which the predicate ‘is wise’ is true. But how can we use these semantic features of the sentence ‘Socrates is wise’ to yield the reference of ‘the proposition that Socrates is wise’? Horwich does not give us any guidance here. Could we say that expressions like ‘the proposition that Socrates is wise’ are semantically unstructured, or at least that after the words ‘the proposition that’ (taken as a functional expression) a sentence becomes a semantically unstructured name of the proposition it expresses? Taking this course would leave us with an infinite primitive vocabulary, and the appearance of the words ‘Socrates is wise’ in two places in the schema would be of no help in understanding the schema or its instances. A further proposal might be to modify our instance of the schema to read:

The proposition expressed by the sentence ‘Socrates is wise’ is true if and only if Socrates is wise.

But following this idea would require relativizing the quoted sentence to a language, a need that Horwich must circumvent.

So let me put my objection briefly as follows: the same sentence appears twice in instances of Horwich’s schema, once after the words ‘the proposition that’, in a context that requires the result to be a singular term, the subject of a predicate, and once as an ordinary sentence. We cannot eliminate this iteration of the same sentence without destroying all appearance of a theory. But we cannot *understand* the result of the iteration unless we can see how to make use of the same semantic features of the repeated sentence in both of its appearances—make use of them in giving the semantics of the schema instances. I do not see how this can be done.

My second difficulty with Horwich’s theory is more dependent on my own further convictions and commitments. Horwich recognizes that to maintain that truth has, as he says, “a certain purity,” he must show that we can understand it fully in isolation from other ideas, and we can understand other ideas in isolation from it. He does not say there are no relations between the concept of truth and other concepts; only that we can understand these concepts independently. There are several crucial

cases so far as I am concerned, since I do not think we can understand meaning or any of the propositional attitudes without the concept of truth. Let me pick one of these: meaning.

Since Horwich thinks of truth as primarily attributable to propositions, he must explain how we can also predicate it of sentences and utterances, and he sees that to explain this without compromising the independence of truth, we must understand meaning without direct appeal to the concept of truth. On this critical matter, Horwich is brief, even laconic. Understanding a sentence, he says, does not *consist* in knowing its truth conditions, though if we understand a sentence we usually *know* its truth conditions. Understanding a sentence, he maintains, consists in knowing its “assertability conditions” (or “proper use”). He grants that these conditions may include that the sentence (or utterance) be true. I confess I do not see how, if truth is an assertability condition, and knowing the assertability conditions *is* understanding, we can understand a sentence without having the concept of truth.

I realize, however, that this is disputed territory, and that heavy thinkers like Michael Dummett, Putnam, and Soames, following various leads suggested by Ludwig Wittgenstein and H. P. Grice, believe that an account of meaning can be made to depend on a notion of assertability or use which does not in turn appeal to the concept of truth.

My hopes lie in the opposite direction: I think the sort of assertion that is linked to understanding already incorporates the concept of truth: we are *justified* in asserting a sentence in the required sense only if we believe the sentence we use to make the assertion is true; and what ultimately ties language to the world is that the conditions that typically cause us to hold sentences true *constitute* the truth conditions, and hence the meanings, of our sentences. This is not the place to argue this. For now I must simply remark that it would be a shame if we had to develop a theory of meaning for a speaker or a language independently of a theory of truth for that speaker or language, since we have at least *some* idea how to formulate a theory of truth, but no serious idea how to formulate a theory of meaning based on a concept of assertability or use.

I conclude that the prospects for a deflationary theory of truth are dim. Its attractions seem to me entirely negative: it avoids, or at least tries to avoid, well-marked dead ends and recognizable pitfalls.

Let me suggest a diagnosis of our aporia about truth. We are still under the spell of the Socratic idea that we must keep asking for the *essence* of an idea, a significant *analysis* in other terms, an answer to the question what *makes* this an act of piety, what *makes* this, or any, utterance, sentence, belief, or proposition true. We still fall for the freshman fallacy that demands that we *define* our terms as a prelude to saying anything further with or about them.

It may seem pointless to make so much of the drive to define truth when it is unclear who is trying to do it: not Tarski, who proves it cannot be done; not Horwich, who disclaims the attempt. Who, then, *admits* to wanting to define the concept of truth? Well, that is right. But. But the same ugly urge to define shows up in the guise of trying to provide a brief criterion, schema, partial but leading hint, in place of a strict definition. Since Tarski, we are leery of the word 'definition' when we are thinking of a concept of truth not relativized to a language, but we have not given up the definitional urge. Thus, I see Horwich's schema on a par *in this regard* with Dummett's notion of justified assertability, Putnam's ideally justified assertability, and the various formulations of correspondence and coherence theories. I see all of them as, if not attempts at definitions in the strict sense, attempts at *substitutes* for definitions. In the case of truth, there is no short substitute.

Now I want to describe what I take to be a fairly radical alternative to the theories I have been discussing and (with unseemly haste) dismissing. What I stress here is the *methodology* I think is required rather than the more detailed account I have given elsewhere. The methodology can be characterized on the negative side by saying it offers no definition of the concept of truth, nor any quasi-definitional clause, axiom schema, or other brief substitute for a definition. The positive proposal is to attempt to trace the connections between the concept of truth and the human attitudes and acts that give it body.

My methodological inspiration comes from finitely axiomatized theories of measurement, or of various sciences, theories that put clear constraints on one or more undefined concepts, and then prove that any model of such a theory has intuitively desired properties—that it is adequate to its designed purpose. Since among the models will be all sorts of configura-

tions of abstract entities, and endless unwanted patterns of empirical events and objects, the theory can be applied to, or tested against, such specific phenomena as mass or temperature only by indicating how the theory is to be applied to the appropriate objects or events. We cannot demand a precise indication of how to do this; finding a useful method for applying the theory is an enterprise that goes along with tampering with the formal theory, and testing its correctness as interpreted.

We are interested in the concept of truth only because there are actual objects and states of the world to which to apply it: utterances, states of belief, inscriptions. If we did not understand what it was for such entities to be true, we would not be able to characterize the contents of these states, objects, and events. So in addition to the formal theory of truth, we must indicate how truth is to be predicated of these empirical phenomena.

Tarski's definitions make no mention of empirical matters, but we are free to ask of such a definition whether it fits the actual practice of some speaker or group of speakers—we may ask whether they speak the language for which truth has been defined. There is nothing about Tarski's definitions that prevents us from treating them in this way except the prejudice that, if something is called a definition, the question of its "correctness" is moot. To put this prejudice to rest, I suggest that we omit the final step in Tarski's definitions, the step that turns his axiomatizations into explicit definitions. We can then in good conscience call the emasculated definition a theory, and accept the truth predicate as undefined. This undefined predicate expresses the *general*, intuitive, concept, applicable to any language, the concept against which we have always surreptitiously tested Tarski's definitions (as he invited us to do, of course).

We know a great deal about how this concept applies to the speech and beliefs and actions of human agents. We use it to interpret their utterances and beliefs by assigning truth conditions to them, and we judge those actions and attitudes by evaluating the likelihood of their truth. The empirical question is how to determine, by observation and induction, what the truth conditions of empirical truth vehicles are. It bears emphasizing: absent this empirical connection, the concept of truth has no application to, or interest for, our mundane concerns, nor, so far as I can see, does it have any content at all.



Consider this analogy: I think of truth as Frank Ramsey thought of probability. He convinced himself, not irrationally, that the concept of probability applies in the first instance to propositional attitudes; it is a measure of degree of belief. He went on to ask himself: How can we make sense of the concept of degree of belief (subjective probability)? Subjective probability is not observable, either by the agent who entertains some proposition with less than total conviction and more than total disbelief, or by others who see and question him. So Ramsey axiomatized the pattern of preferences of an idealized agent who, more or less like the rest of us, adjusts his preferences for the truth of propositions (or states of affairs or events) to accord with his values and beliefs. He stated the conditions on which a pattern of such preferences would be “rational,” and in effect proved that, if these conditions were satisfied, one could reconstruct from the agent’s preferences the relative strengths of that agent’s desires and subjective probabilities. Ramsey did not suppose everyone is perfectly rational in the postulated sense, but he did assume that people are nearly enough so, in the long run, for his theory to give a content to the concept of subjective probability—or probability, as he thought of it.

A brilliant *strategy*! (Whether or not it gives a correct analysis of probability.) The concept of probability—or at least degree of belief—unobservable by the agent who has it and by his watchers, linked to an equally theoretical concept of cardinal utility, or subjective evaluation, and both tied to simple preference by the axiomatic structure. Simple preference in turn provides the crucial empirical basis through its manifestations in actual choice behavior.

We should think of a theory of truth for a speaker in the same way we think of a theory of rational decision: both describe structures we can find, with an allowable degree of fitting and fudging, in the behavior of more or less rational creatures gifted with speech. It is in the fitting and fudging that we give content to the undefined concepts of subjective probability and subjective values—belief and desire, as we briefly call them; and, by way of theories like Tarski’s, to the undefined concept of truth.

A final remark. I have deliberately made the problem of giving empirical content to the concept of truth seem simpler than it is. It would be

*relatively* simple if we could directly observe—take as basic evidence—what people *mean* by what they say. But meaning not only is a more obscure concept than that of truth; it clearly involves it: if you know what an utterance means, you know its truth conditions. The problem is to give *any* propositional attitude a propositional content: belief, desire, intention, meaning.

I therefore see the problem of connecting truth with observable human behavior as inseparable from the problem of assigning contents to all the attitudes, and this seems to me to require a theory that embeds a theory of truth in a larger theory that includes decision theory itself. The result will incorporate the major norms of rationality whose partial realization in the thought and behavior of agents makes those agents intelligible, more or less, to others. If this normative structure is formidably complex, we should take comfort in the fact that the more complex it is, the better our chance of interpreting its manifestations as thought and meaningful speech and intentional action, given only scattered bits of weakly interpreted evidence.

## Notes

1. These words are quoted from Michael Dummett's jacket blurb for Paul Horwich's *Truth* (Cambridge: MIT, 1991). This is not, of course, Dummett's view.
2. I spell out my reasons for rejecting such views in "The Structure and Content of Truth," *Journal of Philosophy*, LXXXVII, 6 (June 1990): 279–328.
3. "The Semantic Conception of Truth," *Philosophy and Phenomenological Research*, IV (1944): 342–360. [Reprinted as chap. 15.—Ed.]
4. "Truth and Proof," *The Scientific American*, CCXX (1969): 63–77.
5. "The Concept of Truth in Formalized Languages," in *Logic, Semantics, Metamathematics* (New York: Oxford, 1956), pp. 152–278 (originally published in German in 1936).
6. "The Establishment of Scientific Semantics," in *Logic, Semantics, Metamathematics*, pp. 401–408.
7. At one time I suggested calling Tarski's concept of truth a correspondence theory on the strength of the role of sequences in satisfying closed sentences, but I subsequently withdrew the suggestion as misleading. For the suggestion, see "True to the Facts," in *Inquiries into Truth and Interpretation* (New York: Oxford, 1984). For the retraction, see "Afterthoughts, 1987," in A. Malichowski, ed., *Reading Rorty* (Cambridge: Blackwell, 1990), pp. 120–138.

8. For references, and further discussion, see my "The Structure and Content of Truth."
9. Cambridge: Harvard, 1961.
10. Cambridge: MIT, 1960.
11. Cambridge: Harvard, 1990. [Reprinted as chap. 20.—Ed.]
12. *From a Logical Point of View*, p. 130.
13. Reprinted in *The Ways of Paradox* (Cambridge: Harvard, 1976).
14. *The Roots of Reference* (La Salle, IL: Open Court, 1974), p. 65.
15. "On the Very Idea of a Third Dogma," in *Theories and Things* (Cambridge: Harvard, 1981), p. 38.
16. "What Is a Theory of Truth?" *Journal of Philosophy*, LXXXI, 8 (August 1984): 411–429. [Reprinted as chap. 17.—Ed.]
17. "Languages and Language," *Minnesota Studies in the Philosophy of Science*, volume VII (Minneapolis: Minnesota UP, 1975).
18. The preceding paragraphs on Quine are partly quoted and partly adapted from a longer and more detailed study of Quine on truth: "Pursuit of the Concept of Truth," in P. Leonardi and M. Santambrogio, eds., *On Quine: New Essays* (New York: Cambridge, 1995). The relevant pages are 7–10.
19. *From Stimulus to Science* (Cambridge: Harvard, 1995), p. 67.

---

## Epistemology and Primitive Truth

Ernest Sosa

Central concepts of epistemology—knowledge, for example, or epistemic justification or intellectual virtue—derive in one way or another from that of truth. Assigning such a central role to truth makes one an externalist in epistemology. What is more, my externalism has a realist cast: the River Nile, the city of Khartoum, and the Mediterranean Sea are so related that the first flows from the second to the third, and does so independently of all beliefs, conventions, cultures, or conceptual schemes. From its meaning alone, the truth of my utterance about the Nile derives from a large and mind-independent feature of African geography. Moreover, when explorers and geographers, astronauts and astronomers, ply their trades, what drives them at least sometimes, at least to some degree, is just plain curiosity—the sheer desire to discover the lay of the land or of the heavens.

### 1 Three Views of Truth

Such a perspective on knowledge and reality comes quickly into conflict with ideas on truth currently on the market. Richard Rorty's new pragmatism, for example, embraces a noncognitivist account of truth, justification, and knowledge. Thus his neo-pragmatist "thinks that the very flexibility of the word 'true' . . . , the fact that it is merely an expression of commendation . . . , ensures its univocity.... So he feels free to use the term 'true' as a general term of commendation in the same way as his realist opponent does . . . , and in particular to use it to commend his own view" (Rorty 1985, 6). For the new pragmatist, there really is no such thing as truth, nor any such thing as justification or knowledge, any more than there is in the world any such thing as hurrayness. The term 'hurrray'

is a functional term that serves for commending and the like; for the pragmatist, the terms 'true', 'justified', and 'known' are like 'hurray'. Thus "for the pragmatist, . . . 'knowledge' is, like 'truth,' simply a compliment paid to the beliefs which we think so well justified that, for the moment, further justification is not needed" (Rorty 1985, 7).

However, our epistemic vocabulary is more versatile than it could be were it merely a vocabulary of commendation. Epistemic terms fit comfortably in linguistic contexts where terms of commendation such as 'hurray' would be out of place.<sup>1</sup>

Another well-known approach explicates a true belief as one that would be accepted rationally and justifiably by ideal inquirers in an ideal epistemic situation for that belief and its subject matter. This sort of approach can be traced back to Charles Sanders Peirce, and more recently has been championed by Michael Dummett, Jürgen Habermas, and Hilary Putnam. Obviously, it is an important alternative that deserves and has received much serious attention. Here I just want to note that if we wish to explain what is rational and justified in terms of what is true, we cannot concurrently explain truth in terms of rationality and justification. Accordingly, such epistemic accounts of truth stay on the shelf, along with neopragmatism, as we move on.

A further sort of account of truth is correspondence to the facts, or correspondence with reality. On this view, the sentences on our lips or in our brains are true when and only when they correspond appropriately with the facts or with reality. Correspondence of a sentence with reality depends, of course, on the reference of its parts. But it is hard to accept such aboutness, intentionality, or reference as a fundamental relation on a par with, let us say, spatiotemporal relations. So we need some acceptable account of reference.

## 2 External Realism

We next consider the prospects for such a correspondence view of truth. Is there a contingent reality external to human minds and language, propertied and interrelated in ways that do not derive from any agreements, stipulations, intentions, thoughts, or experiences on the part of us humans?

*External realism* External reality—e.g., the natural world of hills, streams, mountains, and planets—is not dependent on human thought, and might have been just as it is now even in our absence, and even in the absence of any contingent thinkers whatever, with their sets of categories, beliefs, language, and thoughts of any sort.

In analytic philosophy the attack on this view has intensified notably in recent decades. Indeed, many of the most distinguished and influential thinkers seem to have reached a rare consensus in repudiating realism. Donald Davidson, for example, argues against external realism and correspondence truth, but I will indicate weaknesses in those arguments.

Davidson discusses questions of realism and truth in his Dewey Lectures (1990) and in his paper “Epistemology and Truth” (1987). In these writings, he presents two main arguments against realism:

- (A1) “Neither our beliefs nor our affirmed sentences could possibly be too extensively false, since their having the contents or meanings that they have depends on their being appropriately causally related to the surroundings of the believer or affirmer; which could not happen if they were too extensively false.” (1987, 158–159; compare 1990, 325)
- (A2) Davidson’s own position involves “neither accepting nor rejecting the objectivist-realist slogan that the real and the true are ‘independent of our beliefs.’ The only evident positive sense we can make of this phrase that consorts with the intentions of those who prize it derives from the idea of correspondence, which must be rejected.” (1987, 155; compare 1990, 305)

“The real objection to correspondence theories is . . . that there is nothing interesting or instructive to which true sentences might correspond” (1990, 303). In particular, Davidson has long defended the view that there can be no objectively existing facts to which our sentences might correspond.

We take these two arguments in reverse order. First of all, it is not obvious why the only “evident positive sense” that we can make of the phrase ‘independent of our beliefs’ that “consorts with the intentions of those who prize it” should involve the idea of correspondence.

It might be argued that in supposing that reality might have been just as it is now even in our absence, we must be supposing that certain facts

or states of affairs might have existed even in our absence. And these are just the sorts of entities that, according to Davidson, do not exist. But is the realist committed to accepting such entities? That is far from clear. Perhaps in supposing that reality might have been just as it is now, we suppose only that there are now certain individuals with certain properties and entering into certain relations, such that they might have had these same properties and entered into these same relations even in our absence. (Thus, of course, the Earth could not have been populated by us in our absence, but it might still have been round, etc.)

Unless we opt for substitutional quantification with predicate variables, however, it is hard to see how to state generic realism without such appeal at least to properties and relations. But not even such appeal is needed for recognizably realist positions of a less generic cast. Thus consider the view that there are the nine planets—Mercury, Venus, Earth, Mars, etc.—as well as the Sun, which they orbit; and that these planets might have orbited this sun even in the absence of any humans or thinkers at all to so much as take note of it, much less bring it into being. This is a realism about our planetary system that is modest by refusing to generalize and that can thus avoid quantification over properties (and even substitutional quantification with regard to predicate variables).

What, then, of argument A1? Is this what rules out realism? It is often assumed nowadays that the realist, or at least the *metaphysical* realist, is committed to the view that our total set of beliefs, and even an epistemically ideal set of beliefs, might still have been false. But that depends on an internalist epistemology very much in dispute at present (and for quite a few years already). Cartesian internalists, for whom our epistemic virtue derives solely from our use of reason and introspection, are, of course, committed to the possible falsity of even the most apt sets of beliefs, insofar as they are realists about external reality, which might then conceivably go its own way in complete disregard of our subject's exemplary internal virtue. Thousands of journal pages ago, however, such Cartesianism was already controverted, as were other views of epistemic virtue according to which such virtue depends on the subject's nature and environment in such a way that a subject who is virtuous in a certain environment could not possibly turn out to be globally wrong about that environment.

What is more, there is also a much debated externalist philosophy of mind and psychology according to which a subject could not so much as *have* a rich set of beliefs about her environment if these turn out too extensively false. On this view, we need not even bother with the question of whether these beliefs show epistemic virtue or not. The fact is that they could not even be there as beliefs while being extensively false about the surrounding world. This is what Davidson alludes to with his argument (A1).

My puzzlement about both these forms of externalism—the epistemic and the psychological—is why, given such externalism, it should then be thought that the impossibility of massive falsehood tends to rule out realism. This seems to me like supposing that epiphenomenalism tends to rule out realism. How so? Well, for the epiphenomenalist, there cannot be any mental phenomena except as supervenient on the physical. Accordingly, you cannot have a lot of mentality without that fact entailing a physical reality. But why should it be thought that this rules out a physical reality with the right sort of independence, the sort of independence involved in a realism about external reality? In the first place, even if the sort of physical reality required to yield epiphenomenal mentality could not but be there once the mentality was in place, that is compatible with there being much other physical reality entirely unconnected with any sort of mentality. And anyhow, even in the cases where there is the connection between a sort of physical reality and its corresponding epiphenomenal mentality, the resulting “dependence” of the physical on the mental is not the sort of dependence that rules out realism about the physical, since it does not rule out that *in a relevant sense* the physical still *does not depend* on subjects and their beliefs, categories, conceptual schemes, experiences, and culture. The relevant sense is precisely that the physical is not constituted by the intellectual and cultural activities and mental lives of subjects. For the external realist, the physical does not derive or emerge or depend on our mental lives. On the contrary, there is a complex physical reality constituted by physical individuals such as particles and their properties, relations, magnitudes, forces, etc., and this physical reality is as it is on its own and not in virtue of or derivatively from how it is with our mental or cultural lives. Such external realism is surely compatible with epiphenomenalism. Indeed, on the epiphenomen-



alist view, it is the mental and the cultural that derive from underlying physical reality, rather than the reverse, even though once the mental and cultural are in place, that may entail restrictions on how the physical must be (if it is to underlie the mental and cultural reality and give rise to it). Compare how the temporal array of images on a screen relates to the spatial array of images on the corresponding film.

Nevertheless, correspondence truth still seems problematic. Its main problem at present is not an excess of unacceptable consequences, however, so much as a deficiency of acceptable content. If the truth of sentences—linguistic, mental, or cerebral—is to receive a substantive correspondence analysis, it is indeed plausible that the reference of sentential constituents will be crucially involved, and that such reference will work through some sort of appropriate causal relations. But much work is still required before we have an acceptable understanding of the causal relations involved, and a substantive correspondence theory still seems a distant objective. Accordingly, substantive correspondence truth also stays on the shelf, and we move on.

### 3 Deflationism

We can now more easily sense the attractions of a streamlined deflationism about the concept of truth. For deflationism, truth is not a property that needs a substantial theory, in contrast to the correspondence theory, or a coherence theory, or the traditional pragmatist theory that the true is what works, or the theory that the true is what is validated by one's culture, or any other such theory that tries to ascribe some cognitive content to the concept of truth, or any such theory that considers truth a property whose "nature" needs to be explained.

In a well-known paper, Michael Williams advocates deflationism as follows:

The disquotation schema tells us just about all we need to know about truth. (Williams 1986, 223)

[It is *not* the case that] truth is the name of a property whose nature neither the disquotation schema nor remarks on the use of 'true' to commend etc. so much as hint at, but which theories of truth, properly so called, must try to explain. (Williams 1986, 223)

Why has it been supposed that we need any such substantive property of truth, anyhow? Two main reasons are on offer, according to Williams, and in a way these are really two forms of the same reason. The first is the view that we need to appeal to truth in order to explain the success that we enjoy in negotiating an ordinary day. At every turn, we depend on our beliefs, most of them quite banal and trivially true; fortunately so, since we stake a lot on them, including our survival. We succeed as well as we do partly because we are right as often as we are, that is, partly because we have many true beliefs that serve us well, given our interests and desires. Similarly for our collective success in the use of our technology. Here we succeed as well as we do in part because the scientific theories that underlie such technology are nearly enough and often enough true. To these arguments Williams responds as follows:

To spell out the explanation [of the success of our methods by the truth of our theories] we should have to assert, in a qualified way, all the theories we currently accept, or all those belonging to “mature” sciences. But the predicate ‘true’ saves us the trouble, for we can compress this tedious rehearsal of current views into a single sentence and say “Our methods work because the theories that inform them are approximately true.” (Williams 1986, 230)

If I want a cold drink and believe that the refrigerator, rather than the oven, is the place to look, I will increase the likelihood of getting what I want. This is because cold drinks are more likely to be found in the refrigerator than in the oven. To say that my having true beliefs makes it more likely that I will attain my goals is just a compact way of pointing to an indefinite number of mundane facts of this sort. (Williams 1986, 232)

This is widely regarded as an adequate response to the claim that we need substantive truth in order to explain our individual or collective success, whatever degree of it we enjoy anyhow. I confess to some lingering doubt, based on my uncertain grasp of the relevant operations of “abbreviating,” “compressing,” and “compacting” appealed to.<sup>2</sup> *Stipulative* “abbreviation” is, of course, clear and unobjectionable, but just as clearly, that is not all the “abbreviating” that, on the proposal under discussion, the vocabulary of truth must be expected to do. Even once this term had been adequately explained, what would follow from the fact that a certain vocabulary enabled one to abbreviate in the way suggested? Would it establish that no real properties are expressed by such vocabulary? Is there no real property of the world expressed by “For all

$x, Fx$ ” just because this may be viewed as some sort of abbreviation of “ $Fa_1, Fa_2, \dots$ ”? There would be some irony in the result that universal quantification has no real (“substantive”?) place in scientific explanation, since it is just a way of abbreviating a long conjunction, perhaps an infinite conjunction. Well, if truth turns out to have in explanation a place analogous to that of universal quantification, perhaps that is good enough for any point in the making, whether we end up calling truth a real property or not. Perhaps its status is secure and substantive enough if it is like that of quantification.

A deflationist is not unwilling to define truth, so long as the definition is not of any traditional, “substantive” variety. Some deflationists accept, for example, a substitutionalist paraphrase.<sup>3</sup> Does the sentence “Her statement is true” attribute a “property” to her statement? Through the substitutionalist paraphrase, this question reduces to the question of whether in the following we have such a property: “For all  $p$ , if  $p$  is identical to *her statement*, then  $p$ .” Does the unitalicized part of this sentence express a property attributed to the referent of the italicized singular term? Well, if it expresses *something*, whatever it expresses is true of her statement contingently, and *not* true of many other statements. So it expresses something that is contingently true of some entities and contingently false of other entities. It is this “property,” moreover, that we appeal to when we explain our individual and collective success by saying that enough of our pertinent beliefs and theories have it. And presumably we appeal to this property in supposing to be obviously true every instance of the disquotation schema ( $p$  is true-in-our-language- $L$  iff  $p$ ). The supposed feature that every instance of the disquotation schema is thus said obviously to have is then presumably just the feature attributed to “her statement” above: being such that for all  $p$ , if  $p$  is identical to it, then  $p$ .

#### 4 Varieties of Deflationism

Stepping back, I discern three main varieties of deflationism: (a) non-cognitivism, (b) substitutionalism, and (c) minimalism. I now discuss these in turn.

Both neopragmatism and the redundancy theory are forms of noncognitivism, since both *deny* that the vocabulary of truth has the function of expressing any special property or set of properties, such as correspondence with reality or the like. Instead, truth locutions are like ‘hurray’ in having some other function in our speech—perhaps commending, perhaps emphasizing—and in any case some pragmatic, noncognitive function.

Those who already accept eliminativism of the mental, naturalization of epistemology, and noncognitivism in ethics will find much to commend in such epistemic noncognitivism, oxymoronic though it may seem. Our truth-centered epistemology is incompatible with epistemic noncognitivism, however, since it is committed, for example, to this:

(V)  $x$  is an intellectual virtue only if  $x$  leads mostly to true beliefs.

How is (V) to be stated in harmony with the view that ‘true’ is “merely an expression of commendation,” merely a functional term like ‘hurray’? There seems no way: the role of truth in principle (V) is not just that of an exclamation mark or of a functional term of commendation; it seems rather to be that of a bona fide concept. As is well known, the point may be extended much further. For many other sentences that incorporate the vocabulary of truth pose equally serious problems for noncognitivist neopragmatism. Consider these:

- Some of our beliefs might be false, but many must be true.
- If what they said is not true, they must have lied.
- There are infinitely many truths expressible in English that no one will ever express.

How can we understand the uses in such sentences of the terms ‘true’, ‘false’, and their cognates while abiding by the view that the only function of such vocabulary is to commend, as in the case of ‘hurray’? Neopragmatism is an offer that we can and should refuse.

The redundancy theory for its part holds, naturally enough, that the vocabulary of truth is “redundant” and serves at most for added emphasis, like the exclamation mark, or for some other such noncognitive purpose. Therefore, the redundancy theory falls to the arguments just brought against neopragmatism, as may be seen through obvious

transformations of those arguments. Accordingly, we turn to our second main form of deflationism: substitutionalism.

Substitutionalists are willing to define truth, but their definition is not of any traditional, “substantive” variety. Rather, they offer the following:

$x$  is true  $=_{df}$  there is a proposition, that  $p$ , such that  $[(x = \text{that } p) \ \& \ p]$

And here the quantification in the definiens is “substitutional”: what is required for the truth of the definiens is that we (as issuers of the definition) have in our language a true substitution instance of  $[(x = \text{that } p) \ \& \ p]$ , for some substitution of a declarative sentence of our language in place of the sentential variable ‘ $p$ ’.

The problem for substitutionalism is that what is true always exceeds the bounds of one’s present language: there are always truths not expressible in one’s idiolect of any given time. So there is always some *true* proposition  $x$  such that there is in our idiolect *no* true substitution instance of  $[(x = \text{that } p) \ \& \ p]$ , all of which refutes our substitutionalist definition.

However, not all forms of deflationism are associated with substitutional quantification. An attractive alternative is offered by Paul Horwich, who explicitly rejects the substitutional account, and whose view focuses on the following schema (Horwich 1990):

(1) The proposition that  $p$  is true iff  $p$ .

Anyone who has the concept of truth, notes Horwich, anyone who grasps the property of truth, will accept all propositions expressed by instances of (1) with respect to particular declarative sentences of their idiolect. Horwich’s minimalist theory (MT) is constituted by all such instances of the “equivalence schema” (1) above: “Our theory of truth ... is a collection of propositions—those expressed by instances of [(1)]” (Horwich 1990, 37–38).

We are warned that the theory of truth is irremediably infinite, that “we should not expect to contain all instances of the equivalence schema within a finite formulation: an infinity of axioms is needed. And since this would seem to be an unavoidable feature of any adequate theory of truth, it should not be held against MT. Therefore, we must acknowledge that the theory of truth cannot be explicitly formulated. The best we can do is to give an implicit specification of its basic principles” (Horwich 1990, 31; compare pp. 21, 11–12).

In showing how MT can explain all the phenomena of truth predication that need explaining, Horwich postulates the following: given “the meaning of ‘implies’, we have every proposition of the form  $\langle\langle p \rangle\rangle$  implies  $\langle q \rangle \rightarrow (p \rightarrow q)$ ” (to be read as “(the proposition that  $p$  implies the proposition that  $q$ ) only if ( $p$  only if  $q$ ),” where ‘ $\rightarrow$ ’ is read as the material conditional) (Horwich 1990, 23). Hence, by the meaning of ‘entails’, we must also have the principle schema:

(PE) (the proposition that  $p$  entails the proposition that  $q$ ) only if ( $p$  only if  $q$ )

And, given (PE), why not formulate a finitely minimal theory of truth as follows:

(FMT) Every proposition is necessarily equivalent to (entails and is entailed by) the proposition that it is true.<sup>4</sup>

This theory captures the infinity of axioms constitutive of MT and, obviously, does so through a finite formulation. Or so I will now argue.

We begin with a more symbolic formulation of (FMT):

- (a)  $\langle x \rangle [x \Leftrightarrow \langle x \text{ is true} \rangle]$  by (FMT)
- (b)  $\langle \text{Snow is white} \rangle \Leftrightarrow \langle \langle \text{Snow is white} \rangle \text{ is true} \rangle$  by (a), universal instantiation
- (c) Snow is white iff  $\langle \text{Snow is white} \rangle$  is true by (b), (PE)

And in general, take any arbitrary member of MT, expressed by some arbitrary instance of the equivalence schema (E):

(E)  $\langle p \rangle$  is true iff  $p$ .

We can now derive (E) as follows:

- (a)  $\langle x \rangle [x \Leftrightarrow \langle x \text{ is true} \rangle]$  by (FMT)
- (b)  $\langle p \rangle \Leftrightarrow \langle \langle p \text{ is true} \rangle \text{ is true} \rangle$  by (a), universal instantiation
- (c)  $p$  iff  $\langle p \rangle$  is true by (b), PE
- (d)  $\langle p \rangle$  is true iff  $p$  by (c), truth-functional logic

Therefore, given only the familiar modal notion of entailment among propositions, we can after all capture the infinity of MT within the finite formulation provided by (FMT). At least, we can do so provided we are allowed appeal to (PE).

In any case, for Horwich, minimalism applies not only to propositional truth but also to truth as a property of beliefs and even utterances. Thus he writes, "I shall go on to argue that the minimalist conception applies equally well to the 'truth' of utterances, mental attitudes, and other types of entity" (1990, 7). And he goes on to formulate an interesting "minimalist" theory of truth for utterances, one that is meant to deal with problems raised by indexicals, demonstratives, and ambiguity. But it seems to me that we can best isolate some main philosophical issues if we focus on his first approximation to a minimalism of truth for utterances, as it appears on p. 103: "The initial deflationary impulse is to say that (D?) Any declarative utterance of the sentence-type '*p*', is true iff *p*."

Let us restrict ourselves to sentences that have no indexicals or demonstratives and that are perfectly univocal: call them "eternal sentences." Is (D?) an acceptable minimalism of truth for utterances of eternal sentences?

A problem with (D?) is that it does not give us a general set of necessary and sufficient conditions for the truth of an utterance, nor does it even give us a theory in terms of all the instances of some biconditional schema, as MT does. It only says that *if u is a declarative utterance of a sentence type '*p*', then u is true iff p*. We can overcome this with a two-part theory as follows:

- (D1) Utterance *u* is true iff *u* is an utterance of a (translation of a) sentence type '*p*' that is true.
- (D2) The sentence *Snow is white* is true iff snow is white, the sentence *The sky is blue* is true iff the sky is blue, and so on for all the instances of the equivalence schema (E sent) obtained by filling in with declarative sentences of English or possible extensions of English.
- (E-sent) The sentence *p* is true iff *p*.

Let '(D)' stand for the "conjunction" of (D1) and (D2). A problem with (D) is that it is no longer "pure." It involves a relation between an utterance and a sentence type of that utterance, namely, being an utterance of (a translation of) that sentence type. And this is not just a simple matter of the utterance being of a certain shape, or the like. It is not just a syntactic matter. On the contrary, an utterance in German of 'Schnee ist

weiss' would fit (D) by being an utterance of (a translation of) 'Snow is white', and what makes it so is the fact that it is an utterance of 'Schnee ist weiss' while 'Schnee ist weiss' is translatable as 'Snow is white'. And with this latter relation we get into the matter of linguistic use or the role of an expression in a linguistic community as its members interact with each other and with the surrounding world. So it is hard to see how (D) can be in any appropriate sense "minimalistic." It is very unlike MT in this respect.<sup>5</sup>

## 5 Minimalism and Primitivism

The only form of deflationism that now appears to remain viable is minimalism with regard to the truth of *propositions*, and *not* that of utterances, beliefs, or sentences.<sup>6</sup> But it is not so clear exactly what is involved in this view and what makes it deflationary, even just with regard to *propositional* truth (never mind the truth of sentences, utterances, beliefs, etc.). As I understand it, minimalism says at least two main things:

- (M1) The axioms of the minimal theory are the propositions expressed by instances of the schema (T) 'The proposition that *p* is true iff *p*', for declarative sentences in any possible extension of English. (Call these axioms, collectively, MT. MT is the minimalist theory of truth constituted by axioms each of which is a proposition expressed by an instance of schema (T) for some declarative sentence in some possible extension of English.)
- (M2) One's grasp of the concept of truth, of the property of being true (as applied to propositions), consists in one's disposition to affirm without evidence any member of MT. (And one's understanding of the predicate 'is true' consists in one's disposition to affirm without evidence any instance of schema (T).)

There is, however, an early theory of truth, advocated briefly by Moore and Russell—call it 'primitivism'—that I would like to compare with deflationist minimalism. We are down to these two views as the options open to a truth-centered epistemology. And it seems to me not easy to distinguish clearly between the two, between minimalism and primitivism. Primitivism is described by Moore in the following passage:



It is a theory which I myself formerly held, and which certainly has the advantage that it is very simple. It is simply this. It adopts the supposition that in the case of every belief, true or false, there is a proposition which is what is believed, and which certainly is. But the difference between a true and a false belief, it says, consists simply in this, that where the belief is true the proposition, which is believed, besides the fact that it *is* or “has being” also has another simple unanalyzable property which is possessed by some propositions and not by others. The propositions which don’t possess it, and which therefore we call false, *are* or “have being”—just as much as those which *do*; only they just have *not* got this additional property of being “true.”<sup>7</sup>

What this denies, N.B., is that truth has a Moorean “analysis” or definition (unless we countenance a substitutional account, which would not be a substantive account as traditionally understood). It does not deny that we may know a priori lots of propositions constituted in part by our primitive concept of truth.

When Moore viewed truth as a simple, unanalysable quality, he viewed it just as he viewed *good* in *Principia Ethica*. So he surely would have said of truth what he did say of *good* as follows:

If I am asked “What is good?” my answer is that good is good, and that is the end of the matter. Or if I am asked “How is good to be defined?” my answer is that it cannot be defined, and that is all I have to say about it.... My point is that “good” is a simple notion, just as “yellow” is a simple notion; that, just as you cannot, by any manner of means, explain to anyone who does not already know it, what yellow is, so you cannot explain what good is. Definitions ... which describe the real nature of the object or notion denoted by a word ... are only possible when the object or notion in question is something complex.... But yellow and good, we say, are not complex: they are notions of that simple kind, out of which definitions are composed and with which the power of further defining ceases. (Moore 1903, 6–8)

On this view, what you cannot do with *good* or with *yellow* or with *truth* is to define it, to give an illuminating, compact, at least surveyable, Moorean analysis of it. It is in *this* sense that you cannot philosophically “explain” any such “simple” concept. And this leaves it open that you should have a priori knowledge of infinitely many propositions constituted essentially by such concepts. Among such propositions we may perhaps find the following:

No surface that is yellow all over is blue all over.

Pleasure is intrinsically good.

That snow is white is true iff snow is white.

Let's try to compare now Moore's early primitivism with deflationist minimalism. Primitivism can actually accept both components of minimalism before us: both (M1) and (M2). So there may be more to minimalism beyond (M1) and (M2), and perhaps it is the additional content of minimalism that will distinguish it from primitivism.

Noncognitivist views can explain easily how they are deflationary. For these views, the vocabulary of truth is a functional vocabulary of emphasis or commendation, as is a word like 'hurray'. And just as there is no property of hurrayness, so there is really no property of truth. *This* is clearly deflationary. But this view has unacceptable consequences, and it is not this sort of deflation that minimalism is committed to. For minimalism there *is* a property of truth.

The additional content of minimalism that may explain how it deflates truth and how it is distinguished from primitivism may be found in further minimalist claims about that property, such as (M3):

- (M3) Truth is *not* an ordinary property, and is not a complex or naturalistic property, and neither is reference, and there is no need (or possibility) of a conceptual analysis or definition of truth.

But (M3) *also* can be shared by Moore's primitivism. Compare Moore's views about the property of *good*. So we are still searching for what it is about minimalism that will finally explain how it is deflationary and how it is different from primitivism. And our search next leads us to the following two minimalist claims:

- (M4) Our understanding of truth is pure and independent of other ideas. It is constituted by MT. (Here MT is, again, the minimalist theory of truth consisting of axioms each of which is given by a substitution instance of schema (T): 'The proposition that *p* is true iff *p*'.)

- (M5) There is no theory of truth to be found beyond MT.

Moore's primitivism *can*, I believe, accept even (M4) and (M5). But it needn't. So perhaps it is by sticking its neck out beyond primitivism, by positively accepting (M4) and (M5), that minimalism goes beyond primitivism.

Minimalism so interpreted faces a dilemma. Minimalist claims (M4) and (M5) can be interpreted very modestly, and then they seem eminently defensible, but also compatible with many other theories of truth, including correspondence theories and, of course, Moorean primitivism. On the other hand, (M4) and (M5) can be interpreted less modestly as theses with more content and bite. And in this case they will *not* be compatible with correspondence theories, and anyhow will certainly go well beyond primitivism. But then (M4) and (M5) become very controversial and not at all evident. Let us have a closer look at that dilemma.

Concerning (M4), for example, we might wonder whether our understanding of truth is said to be constituted by MT *wholly* or *partially*. And we might also wonder what it means to say that it is so “constituted.” This opens a number of interpretational possibilities, of which we shall focus on two opposite extremes. On a *weak interpretation*, the claims of (M4) and (M5) are tantamount to the following:

- (M4-W) Our understanding of truth is pure and independent of any other specific idea (such as justification, correspondence, coherence, or the like). It is constituted *in part* by MT in the sense that no one would understand truth adequately unless they were disposed to affirm without evidence each and every member of MT.
- (M5-W) There is no “theory of truth” to be found beyond MT, i.e., there is no *further* axiom *A* such that anyone who understood truth must be disposed to accept *A* without evidence. *Only* the members of MT have this status (apart from simple and obvious *formal* truths involving truth, such as “No proposition is both true and not true”).

On this interpretation, minimalism is a plausible and defensible view. True, there *do* seem to be some other nonformal simple axioms whose claim to being partially constitutive of our understanding of truth is about as good as that of MT, for example, the axioms given by instances of the following schema:

- (KT) *S* knows that *p* only if it is true that *p*.

But this objection seems toothless, since minimalism can respond in either of two ways. First, it can claim that MT actually *can explain* the

supposed additional axioms, in the way it can explain (KT) when MT is combined with instances of the schema (K):

(K) *S* knows that *p* only if *p*.

Alternatively, MT can accept supplementation with a few additional axiom schemata like (KT). So again, on this interpretation, which opts for (M4-W) and (M5-W), minimalism seems plausible and defensible.

However, such weak minimalism is compatible with primitivism and even with strong traditional theories of truth. For example, one could hold a correspondence theory of truth and still accept both (M4-W) and (M5-W), so long as one held the “correspondence theory” only as an a priori and necessary generalization, e.g.,

(C) Necessarily, *x* is true iff *x* corresponds to reality.

So long as one does *not* hold that (C) is true just by meaning, or that one cannot understand truth wholly unless one is disposed to accept (C) without evidence, one can consistently accept both weak minimalism *and* a correspondence account of truth, (C). Such an a priori, necessary “correspondence theory of truth” is compatible with both (M4-W) and (M5-W).

Near the opposite extreme, one could opt for a very strong interpretation of (M4) and (M5), as follows:

(M4-S) Our understanding of truth is pure and independent of any other specific idea (such as justification, correspondence, coherence, or the like). It is constituted *wholly* by MT in the sense that no one would understand truth adequately unless they were disposed to affirm without evidence each and every member of MT, *and there is no further proposition that enjoys this status of the members of MT*.

(M5-S) There is no “theory of truth” to be found beyond MT, i.e., there is no further nonformal necessary biconditional that essentially involves the concept of truth.

*Strong* minimalism, defined in part by (M4-S) and (M5-S) is, of course, incompatible with correspondence theories like (C), obviously so, since C claims in effect that there *is* a necessary biconditional involving truth, one that is not just formal and that does go beyond MT. And strong

minimalism does go well beyond primitivism, since primitivism does *not* make any such strong negative claims as those involved in (M4-S) and (M5-S). Finally, in a very clear sense strong minimalism is deflationary, since it explicitly deflates the claims of traditional theories of truth such as our simple correspondence theory (C).

However, strong minimalism is also less than obvious. Indeed, it is not easy to see how its strong negative theses are to be defended. How do we know that no necessary biconditional such as (C) will ever be found? Before Galileo and Newton, it would have been easy to despair of our ever finding any general laws of motion, and before Euclid and his predecessors, of our ever finding any general geometry. And something similar can be said about Frege and modern logic. It is difficult and implausible to defend strong negative *dicta* such as (M4-S) and (M5-S). Insofar as minimalism does go beyond primitivism and does in a clear sense *deflate* truth, it becomes implausible and debatable.

Minimalism still has a trump card, however, in an appeal to ontological explanation and simplicity. Our disposition to accept MT “constitutes” our understanding of truth, we might now say, by providing an “implicit definition” of the concept of truth, so that there is really no special property of truth beyond what is conveyed by MT. Compare the stipulative definition (B):

(B)  $x$  is a suprateen =<sub>df</sub>  $x$  is a person older than 19.

By this definition a new word is introduced into the language as an abbreviation for a longer expression. Compatibly with this, it may be held that there is here no new property beyond those involved in the definiens, such as the property of being a person, the relation of being older than someone else, etc. Similarly, when we see MT as an implicit definition (not one that is explicit and stipulative, as in (B) above), this enables us, according to minimalism to deny that there is any special property of *truth* beyond properties like being snow, being white, etc.

Such a notion of “implicit definition” seems murky by comparison with the explicit, stipulative sort of definition involved in (B). It is not at all clear when we have an implicit definition that “constitutes” understanding of a predicate. When is a theory to be viewed as such an implicit definition of its constitutive and characteristic terms or concepts? Or are

*all* scientific, mathematical, and philosophical theories to be viewed as thus implicitly definitory of their constitutive notions and terms? If they *all* are, then, of course, much more is deflated than just truth, and we will need a detailed defense of so ambitious a thesis. If only *some* theories are implicit definitions, on the other hand, then we need a way to distinguish those that are from those that are not, and we need an argument to place MT in the former group.

As we have seen, primitivism shares a lot with minimalism, and indeed the two are not easy to distinguish if we interpret minimalism weakly as with (M4-W) and (M5-W). But primitivism does *not* share strong negative theses such as those that define *strong* minimalism, e.g., (M4-S) and (M5-S). Until such theses receive the defense that they require, we do best to opt for the more modest primitivism.<sup>8</sup>

## 6 Conclusion

In conclusion, we may better appreciate the attractions of the following combination of views:

- Truth is a primitive concept and has no illuminating definition or Moorean analysis.
- There is no philosophically substantive theory of truth, one that provides a priori necessary and sufficient conditions for truth in terms of ideal inquirers, pragmatic working, comprehensive coherence, or causal correspondence. Or at least there is no such theory presently on offer.
- We grasp our primitive and unanalyzable concept of truth well enough that we would accept the proposition expressed by *any* instance of the following schema for any declarative sentence in any possible extension of our language: 'The proposition that *p* is true iff *p*'.
- We also have a good enough grasp of our concept of truth to know that when we are curious about some question or field, we want to acquire beliefs regarding that question or that field that are true, and often enough we want to know just to know, just for its intrinsic value.
- Given realism, once everything intentional and semantic is settled, once it is settled what people are believing through their brain or mind or soul states and what people are saying through their utterances, but *without* it being settled which are true and which are not, and once it is settled what reality is nonintentionally and nonsemantically like, i.e., what individuals

exist and how they are propertied and interrelated—once all this is antecedently settled, surely it must follow, as a supervenient necessary consequence of all this, which beliefs or sayings are true and which are not true. (This is more modest than the view that even the semantic and the intentional *also* supervenes on the nonsemantic and nonintentional natural realm. This just says that *truth* is a derivative and contingent property of some beliefs, utterances, etc.)

- It remains to be seen what special shortcomings such a derivative and supervenient concept may have that somehow unsuits it for any interesting explanatory work, unlike the legions of other supervenient and derivative concepts used in science or philosophy or ordinary life.
- Finally, whether we understand truth minimalistically or primitivistically, our grasp of it should be good enough to help shape a truth-centered epistemology by appeal to principles such as (V'):

(V') A belief is epistemically justified or apt iff it manifests an intellectual virtue, a competence to attain the truth and avoid error.

#### Notes

1. The problem was first raised by Peter Geach (1960 and 1965). Among non-cognitivists or fellow travellers who have made serious attempts to deal with the Geach problem, two stand out: Simon Blackburn (1995) and Allan Gibbard (1992). I shall return to epistemic noncognitivism in section 4.

2. In what follows I will abbreviate 'abbreviating, compacting, or compressing' as just 'abbreviating'.

3. See Williams 1986, 234:

The disquotation schema reveals truth as a device for semantic ascent *and descent*: in other words, it suggests that talk of truth can always be paraphrased away. An obvious way to eliminate talk of truth from ... laws linking confirmation by our methods with truth would be to rewrite them as

For any *p*, if *p* is confirmed by our methods then probably *p*.

True, this involves quantifying over propositions or statements. But this is something we might want to do anyway—e.g., to handle sentences like "Everything he said was true."

4. More explicitly, (FMT) reads, "Every proposition *p* is necessarily equivalent to (entails and is entailed by) the *Russellian* proposition that it is true: the *de re*, "concrete" proposition, with respect to *p*, of its being true." For an account of such "Russellian" propositions, see Horwich 1990, 94–95.

5. Compare McGrath 1997a and McGrath 1997b.

6. For my discussion of minimalism, I am drawing on Horwich's work (1990).

7. From Moore 1953, 261. For discussion of the Moore/Russell view, see Cartwright 1987.

8. Except for the early Moore and Russell, whose commitment to the view was fleeting, I know of only one outspoken advocate of the primitiveness of truth. Donald Davidson has long held the view and elaborates his defense of it in Davidson 1990. According to Davidson, you understand a speaker if you have a Tarskian truth theory for his language *L*. Such a theory would be empty if merely a stipulative theory of "truth-in-*L*." We must understand it rather as a theory in terms of a primitive concept of truth, one that applies across languages not only to *L* but also to *L'*, *L''*, etc. We need this concept of truth antecedent to the language of any given speaker under study in order to obtain the data for our Tarskian truth theory of that speaker. For these data will take the form of T-sentences like the following:

'Theaetetus is seated' is true iff Theaetetus is seated.

And these T-sentences make use of our antecedent concept of truth.

Davidson would perhaps be unsympathetic to the streamlined truth theory sketched earlier, which for one thing postulates propositions. But if we do not claim that we have in any sense *defined* truth, which I do not; if we accept the primitiveness of truth, as I do; then there is no reason why we cannot accept a Davidsonian theory of interpretation. We simply have a fuller, and perhaps more credulous, set of views concerning truth. We accept our streamlined theory in terms of propositions, we accept the centrality of truth for epistemology and the principles that go with this idea, and we can accept also the use of our primitive concept of truth for Davidsonian interpretation of each other. Finally, unlike Davidson, we see no threat to external realism in such a view of truth as primitive and indefinable, even when this is combined with a metaphysical theory of truth as correspondence with reality, external reality included. So we may remain agnostic about the prospects for such a correspondence theory.

The present text draws from earlier publications, including Sosa 1993a and Sosa 1993b.

## References

- Blackburn, Simon. 1995. *Essays in Quasi-Realism*. Oxford: Oxford University Press.
- Cartwright, Richard. 1987. "A Neglected Theory of Truth." In his *Philosophical Essays*, pp. 71–95. Cambridge: MIT Press.
- Davidson, Donald. 1990. "The Structure and Content of Truth." *Journal of Philosophy* 87: 279–328.
- Davidson, Donald. 1987. "Epistemology and Truth." *Proceedings of the Congreso Internacional Extraordinario de Filosofía*, pp. 147–161. Córdoba, Argentina: Universidad Nacional de Córdoba.
- Geach, Peter. 1960. "Ascriptivism." *Philosophical Review* 69: 121–125.



- Geach, Peter. 1965. "Assertion." *Philosophical Review* 74: 449–465.
- Gibbard, Allan. 1992. *Wise Choices, Apt Feelings*. Cambridge: Harvard University Press.
- Horwich, Paul. 1990. *Truth*. Oxford: Basil Blackwell.
- McGrath, Matthew. 1997a. "Weak Deflationism." *Mind* 106: 69–98.
- McGrath, Matthew. 1997b. "Reply to Kovach," *Mind* 106: 581–587.
- Moore, George Edward. 1903. *Principia Ethica*. Cambridge: Cambridge University Press.
- Moore, George Edward. 1953. *Some Main Problems of Philosophy*. London: Allen and Unwin.
- Rorty, Richard. 1985. "Solidarity or Objectivity?" In *Post-Analytic Philosophy*, ed. by J. Rajchman and C. West. New York: Columbia University Press.
- Sosa, Ernest. 1993a. "The Truth of Modest Realism." *Philosophical Issues* 3: 177–195.
- Sosa, Ernest. 1993b. "Epistemology, Realism, and Truth." *Philosophical Perspectives* 7: 1–16.
- Williams, Michael. 1986. "Do We (Epistemologists) Need a Theory of Truth." *Philosophical Topics* 14.

---

## Truth: The Identity Theory

Jennifer Hornsby

I want to promote what I shall call (unoriginally, and for the sake of its having a name<sup>1</sup>) ‘the identity theory of truth’. I suggest that other accounts put forward as theories of truth are genuine rivals to it, but are unacceptable.

A certain conception of *thinkables* belongs with the identity theory’s conception of *truth*. I introduce these conceptions in Part I by reference to John McDowell’s *Mind and World*; and I show why they have a place in an identity theory, which I introduce by reference to Frege. In Part II, I elaborate on the conception of thinkables; and by adverting to interpretive accounts of speakers, I introduce a perspective from which the identity theory’s merits can be revealed.

### I

#### I.1

McDowell introduced the notion of a *thinkable* in order to fend off a particular objection to the following claim (1994, p. 27).

[T]here is no ontological gap between the sort of thing one can ... think, and the sort of thing that can be the case. When one thinks truly, what one thinks *is* what is the case. ... [T]here is no gap between thought, as such, and the world.

Someone who objects to this supposes that, by denying any gap between thought and the world, one commits oneself to a sort of idealism. But such an objector confuses people’s thinkings of things with the contents of their thoughts. If one says that there is no ontological gap between thoughts and what is the case, meaning by ‘thoughts’ cognitive activity on the part of beings such as ourselves, then one is indeed committed to a

sort of idealism: one has to allow that nothing would be the case unless there were cognitive activity—that there could not be a mindless world. But someone who means by ‘thoughts’ the contents of such activity, and who denies a gap between thoughts and what is the case, suggests only that what someone thinks can be the case.

[T]o say that there is no gap between thought, as such, and the world, is just to dress up a truism in high-flown language. All the point comes to is that one can think, for instance, *that spring has begun*, and that the very same thing, that spring has begun, can be the case. That is truistic, and it cannot embody something metaphysically contentious....

In order to avoid the ambiguity in ‘thought’ which would be exploited if a metaphysically contentious idealism were reached, McDowell suggests using the word ‘thinkables’ for what may be thought. My policy here will be to use the word ‘thinkable’ generally, in place of any of the more familiar ‘content’, ‘proposition’ or ‘Thought’. Further reasons for this choice of word will show up in due course.

McDowell’s demonstration that his position avoids a simple idealism may strike some people as an inadequate defence. I think that it can help to defend it to locate it by reference to debates about truth. One may view the quotations from McDowell as encouraging an identity theory of truth.<sup>2</sup> This says that true thinkables are the same as facts. True thinkables then make up the world of which McDowell speaks when he dresses up a truism. The world is ‘everything that is the case’, or ‘a constellation of facts’, as McDowell puts it, following Wittgenstein.

## I.2

The identity theory is encapsulated in the simple statement that true thinkables are the same as facts. But it may be wondered how that statement could amount to a *theory* of truth: ‘If someone asks what truth is, and receives an answer which helps itself to the idea of a fact, then hasn’t she been taken round a very small circle?’ Yes. But the simple statement on its own is not supposed to tell us anything illuminating. A conception of truth can be drawn out from an elaboration of what the simple statement can remind us of. And, as we shall see, the conception can be set apart from the conceptions of other accounts that go by the name of theories of truth.

The identity theory is not vacuous. It cannot be vacuous because it takes a stand on what the bearers of truth are, calling them thinkables. This is not an uncontentious stand. For there are philosophers who have told us that the notion of proposition (and thus of thinkable) is so dubious that we should take the truth-bearers to be sentences.<sup>3</sup> The identity theory proceeds without such doubts, taking it for granted that we can make adequate sense of what is meant when someone says, for instance, 'She told me something that isn't true'.<sup>4</sup> And the identity theory not only asks us to understand such 'something's in appreciating where truth is applicable, but it also asks us to understand such 'something's in saying what truth's applicability consists in. Certainly there is no illumination at the point at which the word 'fact' is resorted to in order to say what this applicability consists in. But the identity theory makes definite commitments nonetheless.<sup>5</sup>

### I.3

Whether or not its title to be a theory can be made out, it may be unclear why the word 'identity' belongs in it. What could be the point in saying that true thinkables *are the same as* facts, rather than—more simply and apparently to the same effect—that true thinkables *are* facts?<sup>6</sup>

A familiar argument in Frege (1918) may help to show the point. It is an argument against the correspondence theory of truth. Frege introduces it with the words 'It might be supposed . . . that truth consists in the correspondence of a picture with what it depicts'. 'This is contradicted, however', he says, and then argues by *reductio* (pp. 18–19):

A correspondence . . . can only be perfect if the corresponding things coincide and so just are not different things at all. . . . [I]f the first did correspond perfectly with the second, they would coincide. But this is not at all what people intend when they define truth as the correspondence of an idea with something real. For in this case it is essential precisely that the reality shall be distinct from the idea. But then there can be no complete correspondence, no complete truth. So nothing at all would be true; for what is only half true is untrue.

Putting this only slightly differently, we hear Frege saying: if truth were explicated in terms of any relation, it would have to be identity, since anything less than a candidate for truth's coincidence with a putatively corresponding thing would lead to the intolerable conclusion that there is

no truth. Someone who takes herself to think that true thinkables correspond to the facts has it right, then, only if she actually means that any true thinkable is the same as some fact—which is what the identity theorist says.

Frege's argument has a sequel. This starts by showing how Frege thinks his opponent will respond. The opponent asks (p. 19):

But can't it be laid down that truth exists where there is correspondence in a certain respect?

Here it is conceded that truth cannot be unspecified correspondence, so to speak. The problem with taking truth to be unspecified correspondence is that there can be correspondence in this respect, or that respect, or that other respect, so that there can be less or more correspondence according as there is correspondence in fewer or more respects; but there can't *in any analogous way* be more or less truth.<sup>7</sup> The opponent supposes that he can get out of this difficulty by picking on one respect of correspondence. To this Frege has a response.

But in which [respect]? What would we then have to do to decide whether something were true? We should have to inquire whether an idea and a reality, perhaps, corresponded in a laid-down respect. And then we should have to confront a question of the same kind, and the game would begin again. So the attempt to explain truth as correspondence collapses.

If there was something distinct from a thinkable (a reality, say) such that establishing that some relation obtained between it and the thinkable was a way of getting to know whether the thinkable was true, then someone could be in the position of knowing what is known when the thinkable is known, yet of still not knowing whether it was true. But of course one could never be in that position: to discover whether *p* is already to discover whether it is true that *p*.

This reveals a general difficulty about defining truth—the difficulty which shows up 'when we confront the same question again'.

In a definition certain characteristics would have to be stated. And in application to any particular case the question would always arise whether it were true that the characteristics were present.

'Consequently', Frege concludes, 'it is probable that the word 'true' is unique and indefinable' (p. 19).

When one follows Frege's argument through to this general conclusion, about the definability of truth, explicit opposition to the correspondence theory is lost: the correspondence theorist's definition fails to meet a constraint on any adequate definition; but it turns out not to be alone in that failure. Frege accordingly might be thought to have argued against an especially naive correspondence theory in the first instance, and then turned to opposing the whole idea of truth's definability. But there can be a point in thinking of Frege's initial argument as meant to show that a correspondence theory in particular—and *any* correspondence theory—is untenable. This is an argument which is sound only if the identity theory escapes its *reductio*. It is the initial argument whose conclusion can be dressed up in high-flown language: there cannot be an ontological gap between thought ('an idea') and the world ('something real'). Given the sequel to the initial argument, the high-flown language can hardly point us toward any substantial theory of truth: truth's indefinability prevents us from thinking that truth has a nature that a theory could spell out. The high-flown language, then, serves only to remind us that a metaphysical stand is taken when an identity theory is endorsed.

#### I.4

The identity theory, at any rate, is distinguishable from any correspondence theory. And the identity theory is worth considering to the extent to which correspondence theories are worth avoiding. I think that correspondence theories *need* to be avoided. I mean by this not merely that they are incorrect, but that people are apt to believe them.

It is common for philosophers to speak as if a correspondence theory of truth had no metaphysical import whatever. We are sometimes told that the idea of correspondence is recorded in a series of platitudes that any theorist of truth has to respect. Simon Blackburn has spoken of the phrase 'corresponds to the facts' as sometimes a piece of Pentagonese—a paraphrase of 'is true' deployed with the purpose of saying something important sounding (1984, p. 255). But of course this is not all that has ever been read into the phrase. Someone who says 're-rendered it operational' for 'got it going again', may be criticized for needless portentousness, but not on other grounds; but when 'corresponds to the facts' gets in, the phrase's wordiness should not be the only source of doubt.

Certainly there are glosses on 'is true' that are platitudinous: 'is a fact' is one such—the one that the identity theory singles out for attention. Perhaps it is also a platitude that true sentences say how things are. And this again is unobjectionable, so long as the 'things' in question are ordinary objects of reference: the true sentence 'that book is red', for example, says something about how things are by saying how one of the things (sc. that book) is (sc. red). This platitude then points up the independence of thinking from what there is. Whether you want to know the book's colour, or to know something of what I think about the book, you have to think of something that is not sustained in existence by your thinking. Still, the thing to which you are then related (that book) is obviously not a correspondence theorist's candidate for the correspondent of a truth-bearer.<sup>8</sup> The platitudes about truth do not record the correspondence theorist's claims about it.

From the point of view introduced by the identity theory, it will be distinctive of correspondence theorists to seek items located outside the realm of thinkables, and outside the realm of ordinary objects of reference, but related, some of them, to whole thinkables. The idea is widespread, and it takes various guises. In the Russell of *An Inquiry into Meaning and Truth* (1940), the basic correspondents are percepts. Percepts can be 'surveyed but not defined'; utterances appropriately associated with them get their particular meanings from them; and propositions, the truth bearers, can be constructed out of percepts. In the Quine of *Philosophy of Logic*, the correspondents are cosmic distributions of particles. 'Two sentences agree in objective information, and so express the same proposition, when every cosmic distribution of particles over space-time that would make either sentence true would make the other true as well' (1970, p. 4). These very different candidates for things that make sentences true—percepts and particle distributions—reflect the very different obsessions of Russell and Quine, epistemological and cosmic. But what is common to their accounts, despite this vast difference, is a willingness to reconstruct thinkables from posited entities of a different sort, entities which make things true. Percepts and particle distributions, then, are supposed to be items which we can specify independently of an account of thinkables, items which may confer truth upon a thinkable. When they

are introduced, however, we cannot hold onto the truism that inspires the identity theory. The fact (as it is) that autumn has begun, if it were to be a cosmic distribution of particles, would not be the same as what I think when I think (truly) that autumn has begun.

It is evident now that the words 'corresponds with' do not have to be in play for an ontological gap between thought and the world to open up. This is something that we see in formulations used over the years by Michael Dummett and Crispin Wright in stating the semantic anti-realist's case. Their formulations often appear to invoke a conception of a truth-maker which will suit a correspondence theorist but which an identity theorist cannot allow.<sup>9</sup> Dummett asked 'If it were impossible to know the truth of some true statement, how could there be anything which *made* that statement true?'. Wright spoke of 'a truth-conferrer for a sentence': in the case where the truth of the sentence cannot be known, he said that this is something that 'the world fails to deliver up'. And he spoke of 'the states of affairs' that are in question when a sentence is undecidable as things that 'could not be encountered'. These ways of speaking give rise to an image of something with which a thinkable might have connected up, but a something which we are expected to think of the world as taking sole responsibility for. This is the image that an identity theory may help to rid us of. For when the conditions for the truth of a sentence are supplied by an identity theorist, nothing is brought in besides the thinkable that is expressed by the sentence itself. By introducing 'sources of truth', 'truth conferrers' and 'states of affairs', Dummett and Wright drive a wedge between what is demanded by a thinkable and what is demanded by a thinkable that is true. The identity theorist leaves no room for any wedge at this point.

Of course these remarks about Dummett and Wright do not get to grips with the position which was their concern. But they can illustrate a point—that philosophers' formulations are apt to create an outlook which is forsworn when an identity theory displaces a correspondence theory. I hope that they also suggest how the identity theory may displace forms of anti-realism more subtle than the crass idealism which results from equating thinkables with thinkings of them.



## II

### II.1

It would be laborious to attempt to show that the identity theory is incompatible with all things irrealist. In order to show that it embodies nothing metaphysically contentious, I shall attempt only to reveal its actual compatibility with a perfectly commonsense realism.

McDowell's rebuttal of any simple idealism emphasizes the independence of thinkables from thinkings. One way to grasp this independence is to see that there are (so to speak) more thinkables than there are thinkings. I suspect that those who find the theory problematic are apt to suppose that it could be part of commonsense that there are (so to speak) more facts than there are true thinkables. If this is right about where the opposition lies, then further reflections on the identity theory, if they are to serve as a defence, must expand on the notion of a thinkable. By the identity theorist's lights, our grasp of the notion of a fact cannot exceed our grasp of the notion of a true thinkable. But someone who wishes to express doubts on that score might be helped by having it made apparent how generous the notion of a thinkable nonetheless is.

### II.2

There can seem to be an immediate obstacle, however, to *any* account of thinkables—of the contents, the meaningful things that bear truth. Quine's attack on the Myth of the Museum is directed against the assumption that there could be things external to thought and meaning, lodged like exhibits in the mind, whose relations to other things could constitute the foundations of meaning (1960). The identity theorist agrees with Quine about the incoherence of the hope that intersubjective sameness of meaning might be explained in terms of relations with things external to thought and meaning.

From the identity theorist's point of view the correspondents of correspondence theories of truth play the same role as the exhibits in the museum of the mind: they are items located beyond the bounds of human play with concepts, in terms of which one is supposed to explain meaning. Quine, speaking of cosmic distributions of particles over space-time, said that the item assigned to one sentence as a condition of its truth is

the same as the item assigned to another sentence as a condition of *its* truth if and only if the two sentences have the same meaning. But such items as cosmic distributions of particles are in the same boat as items in the mind's museum according to the identity theory: neither can be used in the reconstruction of thinkables from something else.

If one countenances the cosmic items, but is led by the problems of the items in the mind's museum to think that ordinary talk of meaning is unsupported, then one may invoke a double standard. Quine tells us that a second class standard is appropriate so long as we are tolerant of such everyday psychological talk as involves any notion of a thinkable (1960, §45). But he said that we can, and in science we must, employ a first class standard; it is then that objective information, corresponding to (say) cosmic distributions of particles, can do duty for thinkables, Quine thinks. The upshot of this is hard to make coherent. For the view of everyday reports of people's psychological states which is required by Quine's lower standard for them is not a view that can be sustained by someone who takes herself (for instance) to seek the truth in some area. A person's being an enquirer of any sort requires that she be interpretable as aiming at gleaning the facts, and we have no conception of what that is excepting as we can think of her as more generally intelligible—as apt to perceive things, and to think them, and to draw conclusions. We cannot then be in a position to make statements about Quine's first-class reality but of refusing (according to the same standard) to make any statements which say, for instance, what people are doing when they are investigating that reality. The identity theory helps to make this difficulty with the Quinean picture vivid. The first-class standard was meant to be the standard of genuine facts; the second-class standard was to be invoked when the language of thinkables was used. But if any fact is the same as some true thinkable, then we cannot endorse facts and despise thinkables.

### II.3

It can seem as though the identity theorist had nowhere to turn for an account of thinkables. At least there is nowhere to turn for an account besides an investigation of other predications to them—predications other than 'is true'. This brings me to further reasons (which I said I would come to) for using the term 'thinkable'.

'Thinkable' is a word for a sort of things to which a person can be related in various modes. I say that the Labour Party will win the next election. I have just said something (that Labour will win) which many now believe, which a good few hope, which John Major fears. The example then shows that thinkables can be beliefs, hopes and fears. They are called beliefs when thought of in connection with one psychological attitude towards them; they are called hopes or fears when thought of in connection with other attitudes. They are thought of as propositions when thought of as propounded.<sup>10</sup> A modal term, like 'thinkable', may serve to remind one of the variety of relations here: it is not only thought which relates to *thinkables*, because a thinkable can be believed and hoped, for instance. (And just as we must not confuse a thinkable with a thinking, so we must not confuse a thinkable with someone's believing one, or with someone's hoping one.)

Besides '\_\_\_ is true', then, there are predicates of thinkables, such as '\_\_\_ is believed by Tony', '\_\_\_ is hoped by members of the crowd'. Yet other predicates of thinkables show people as related to them by their speech acts: a statement, for instance, is what we call a thinkable when we think of it in connection with someone's making a statement. 'Thinkable' gives a word for what is truth-evaluable which is indifferent between the case where the evaluable thing is presented as the object of a state of a thinker's mind and the case where it is presented as having been put into words. But it is the linguistic expression of thinkables which we are bound to focus on, if we are to find anything of a systematic sort to say about them. One aim of theories of meaning is to show the significance of sentences as systematically dependent on properties of the words that make them up: theories of meaning, one might say, treat of thinkables' composition. The productivity of language, which can be revealed in its theory of meaning, then points towards another reason for using a modal notion, and speaking of *thinkables*. Someone in possession of a theory of meaning for some language can say what was expressed in the use of any of the sentences on some list, composed from some stock of words; and is in a position to see that there are other things that would be expressed in the use of other sentences, not on the list, but composed only from words in the same stock. A theory of meaning, though its data are uses of actual sentences, is a theory which speaks to potential uses—

to what would be said if some hitherto unused sentence of the language were used. There are actually unused sentences, which, just like the sentences we have given voice to or heard or read, express thinkables.

This suggests the place to look if we want to expand on the notion of a thinkable. We cannot postulate meanings in the mind or correspondents in the world. But we can look to the actual practices of language users. And we shall be reminded here of an idea first recommended by Donald Davidson—that we might put to work, as a theory of meaning of the language of some speakers, a definition of truth for the language which enables the interpretation of those speakers. Davidson's claim that a definition of truth for a language can serve as its theory of meaning depended in part on his thinking that Tarski had shown a way of displaying the recurrent significance of words—by treating words as having characteristics which affect the truth of sentences they come into.<sup>11</sup> In the present context, much of the importance of the idea of deploying such a definition of truth for a language is the view of predications of thinkables it affords. Where an account of a language's workings is interpretive of its speakers, it enables the theorist to give expression, in the case of any sentence in the language and any speaker of it, to the thinkable expressed by the speaker using that sentence. It thus gives the theorist the resources to say what speakers are doing when they use their language.<sup>12</sup>

An interpretive account of speakers is not narrowly linguistic. For speakers' productions of sentences cannot be seen as intelligible expressions of thinkables except as speakers are seen to have some purpose in producing the sentences. And any hypothesis about the purpose of a person who uses words on some occasion goes hand in hand not only with a hypothesis about the thinkable then expressed but also with hypotheses about her mental states—about how *belief* and *desire* and the other attitudes relate her to thinkables—and with hypotheses also about the states of mind of audiences to her speech, and of all the others who use the language on other occasions.

The imaginary theorist, who compiles the facts about words that could put one in a position to understand foreign speakers, would be involved not only in making attributions to speakers of psychological attitudes and speech acts towards thinkables, but also, and inevitably, in taking a view of the truth of the thinkables to which speakers are then taken to be

related. One cannot generally take a view about what someone's purposes are without having some view of which of those purpose are achieved; people intentionally do what they try to do to the extent that the beliefs which explain their doing what they do are true (are believings of true thinkables, that is). Of course the word 'true' does not have to be dragged in in order to see someone's taking an attitude towards a thinkable as working as it does. One can just as well say 'She believed that the plane took off at 9, and the plane took off at 9' as one can say 'She believes that the plane took off at 9 and that is true'. But insofar as an interpretive account requires more than the idea of people's relations to thinkables, and more than the idea of interconnections between those relations, it requires grasp of the distinction involved in assessments of thinkables as true or false. The view of thinkables that emerges, then, in trying to expand on the notion, is one in which some thinkables are taken to be (the same as) facts.

The study of interpretive accounts affords a distinctive perspective on the application of 'is true' to thinkables. 'True' can be treated as having a role alongside a variety of psychological predicates; but it is not itself treated as a psychological predicate, of course.<sup>13</sup>

## II.4

Discussions of coming to understand a foreign language sometimes assume its speakers to be more ignorant than the theorist: the facts at the theorist's disposal go beyond any of which the interpreted people are apprised. But this assumption is not essential to the idea of an interpretive account. Contemplating interpretive accounts shows the acceptability of a conception of potential uses of language expressive of thinkables outside one's ken, and some of which are facts.

One might think inductively here. Over the centuries, human knowledge, at least in some spheres, has expanded, and its expansion has been assisted by the introduction of new concepts, for instance in the formulation of scientific theories. If one believes that human knowledge will continue to expand, one is entitled to predict that thinkables which none of us here and now is capable of thinking will come to be known. One may envisage a theorist interpreting a language of the future: its speakers would think things, and the theorist, in coming to understand them,

would learn from them. She could come to have access to facts, which in her present situation she is not even equipped to express.

Here one thinks of thinkables in connection with expanding knowledge. And it might then be supposed that the facts are to be circumscribed by reference to what is known by an ideal knower, at the limit, as it were, of an inductive series of more and more knowledgeable beings. But acceptance of unthought thinkables, some of which are facts, requires no such supposition. The supposition requires an understanding of the ideal situation for arriving at knowledge. And this can only be a situation in which all sources of error are eliminated or taken account of—a situation, that is to say, in which one is sure to believe what is *true*. Perhaps we can gesture towards such an ideal. But since we can explain it at best in terms of an antecedent notion of truth, the style of thinking used here to uncover a conception of facts can lend no support to an epistemic theory of truth.<sup>14</sup>

The conception of unthought thinkables elicited here does not depend upon any settled opinion about human ambitions or limitations, but only upon an idea of intelligible others from whom one could learn. It evidently yields a generous conception of facts, to which an identity theorist is entitled. I hope, then, that the identity theory emerges as a defensible theory of truth, in keeping with our commonsensically realist view about the extent of facts independent of us.<sup>15</sup>

Answers to philosophers' questions about the relation between language and the world have traditionally taken a form that we now call theories of truth. I have not meant to develop any new theory here. Indeed, I do not think that we need a theory of truth, save insofar as we may go astray without one. I have promoted the identity theory because I think that we have to find a position from which to avoid the false dilemmas that theories currently on offer present us with. Nowadays many suppose that philosophers *either* endorse some version of a correspondence theory *or* have to say distinctively deflationary things about truth.<sup>16</sup> I hope that reflection on the identity theory shows that dissension from correspondence theories, and indeed from all theories that purport to analyze truth, is independent of a deflationary attitude toward truth. The identity theory, by prompting questions about the nature of thinkables, provides us with a perspective from which many other theories appear indefensible.<sup>17</sup>

## Notes

1. For 'the identity theory' in recent and contemporary philosophy, see Candlish 1995.

2. I do not say that McDowell himself would see a point in viewing them thus.

3. The doubts are induced by Quine's attack on propositions, which I touch on below (at §II.2, and see nn. 6 & 10).

I think that someone who had never encountered logic or semantics might have encountered predications of truth to thinkables without encountering predications of truth to sentences; and the question *what truth is* surely concerns a concept which might feature in a language about which logicians and semanticists had never had anything to say. At a minimum, then, a philosopher who takes *truth* primarily as a property of sentences must say something about what appear to be its predication to thinkables. Although I accord priority to thinkables' truth here, I acknowledge that, when returning answers to particular philosophical questions, the application of 'true' to sentences is indispensable: see below, Part II. I acknowledge also that what appear to be predications of truth to thinkables may be treated as no such thing, as in the prosentential theory (see n. 6). *Pro hac vice* I talk as if the surface appearances were sustainable.

4. In saying that the identity theorist proceeds without doubts, I do not deny that hard work has to be done to give accounts of what appears to be talk about propositions/thinkables. An identity theory of truth evidently places constraints on such accounts. See e.g. Rumfitt 1993's account of the construction of propositions: Rumfitt's constructionalism goes hand in hand with a paratactic treatment of the logical form of sentences containing 'that'-clauses; but his kind of constructionalism might be entertained outside the context of such treatment.

5. Candlish says, of what he calls a 'modest' identity theory, that it is 'completely uninteresting—trivial ... precisely because it has no independent conception of a fact to give content to the identity claim' (1995, p. 107). Candlish assesses the theory as if it had the ambitions of a definition. But what I call 'the identity theory' has no such ambitions; its interest derives from what it can be seen, from what it says, to be opposed to philosophically. Candlish allows that an *immodest* ('robust') identity theory might be interesting: its interest could derive from its 'independent conception of facts', independent, that is, of the conception of thinkables, or truth-bearers. For my own part, I cannot see a point in thinking that such a theory deserves the name of *identity* theory. (Here I disagree with Julian Dodd 1995, from whom Candlish takes the robust/modest distinction. There is much about which Dodd and I agree, however: see our 1992.) Addition to note, 1999: In the three discussion papers in *Proceedings of the Aristotelian Society* 1999, Part 2, pp. 225–245, Dodd (among other things) says why he considers my position to be that of a robust theorist, Candlish (among other things) contrasts my position with Dodd's, and I respond to Dodd and Candlish. See also Dodd 2000.

6. The introduction of 'identity' might seem to have the consequence of upping the ontological stakes (so that thinkables are to be treated as OBJECTS). That is

not so. When we have understood, for example, ‘She does it in one way, and he does it in another way’, we have also made sense of ‘They don’t do it the *same* way’—but not at the expense of treating either things that are done or ways of doing them as OBJECTS. I think that hostility to propositions derives partly from Quine’s assumption that all quantification is objectual or (in Quine’s own sense) substitutional. This assumption has seemed to have the consequence that unless we give a Quinean substitutional account of these ‘something’s, we shall be forced to treat propositions as OBJECTS, in a sense of the term caught up with a particular understanding of singular reference. But Quine’s assumption is not compulsory: see e.g. Davies 1981, Ch. VI, §3. Some of the interest of the prosentential theory of truth, defended in Grover 1992 and Brandom 1994, derives from the directness of its challenge to Quine’s assumption.

The identity theory is not formulated in order to take a stand on the logical form of predications of truth. If taken to reveal logical form, it would take an erroneous stand—the one which is contradicted by Frege’s remark that “‘true” is not a relative term’. Comparison with Russell’s Theory of Descriptions may be helpful here. In the analysis of ‘the’ provided by Russell, the word ‘the’ is not treated as the simple quantifier which, presumably, so far as logical form is concerned, it is. One point of giving the analysis which Russell’s theory states is to show what is involved in seeing ‘the’ as a quantifier, and to show which quantifier it is. Something analogous goes on when ‘identity’ is introduced into an account of truth. Just as Russell’s Theory can present the negative semantical claim that ‘the’ does not combine with predicates to form names, so the identity theory of truth can present its own negative metaphysical claims—claims such as emerge from seeing how the identity theory arises out of rejection of a correspondence theory.

One point of a formulation including ‘same’ might be to draw attention to the principles of distinctness of facts presupposed to the theory: those principles cannot allow a coarser grain to facts than to thinkables. (This means that it is not a target of the so-called slingshot argument; see Neale 1995.) A naive account of facts, attractive to those who seek facts in line with a correspondence conception, might incorporate the principle: Where  $a = b$ , ‘ $Fb$ ’ does not express a different fact from ‘ $Fa$ ’. Such a principle, obviously, is at odds with the identity theorist’s conception of facts. (In Neale’s terms: ‘the fact that ( ) = the fact that ( )’ is –PSST.)

7. Frege pointed out that ‘with every property of a thing is joined a property of a [thinkable], namely that of truth’ (p. 20). For illustration, suppose that Fred is tall. Putting it in Frege’s way, a property of Fred (being tall) is joined to a property of a thinkable: if Fred is indeed tall, then a true thinkable is put forward when Fred is said to be tall. But if this is correct, then it can seem that we should allow that *truth* can have any of the features which the property of *being tall* can have, so that if *being tall* admits of degrees (if  $x$  can be to some extent tall), then *truth* admits of degrees (it can be to some extent true that  $x$  is tall). But now it seems that Frege appreciates a characteristic of ‘true’ which ensures that, when treated as a predicate, it will seem to admit of degrees, if any does. This makes me think that when Frege invokes the claim that what is half-true is untrue, he is



relying on the thought that any *relation* introduced to account for truth cannot be a relation which admits of degrees. And that is why I say that there cannot *in any analogous way* be more or less truth.

8. Davidson used to say that a relation like Tarskian *satisfaction* could provide the language-world links sought by a correspondence theorist of truth. But Davidson now regards this as a mistake (1990, p. 302). It must indeed be a mistake if opposition to correspondence theories can be combined with thought about mind-independent objects.

9. See Dummett 1976, at p. 61 in the version reprinted in Dummett 1993.

A different sort of illustration may be got from Frank Jackson, Graham Oppy and Michael Smith 1994. They argue for the compatibility of versions of non-cognitivism (in ethics, say) with minimalism about truth. They follow Michael Devitt in characterizing minimalism as holding that ‘terms for truth and falsity are linguistic devices for talking about reality by appending the truth predicate’. Their claim then is that it might not be that any old sentence is such as to talk about reality: non-cognitivists, they say, ‘precisely deny that (e.g.) ethical sentences talk about reality’. But someone who is opposed to correspondence theories in all their versions will not allow this ‘talking about reality’. Suppose that Devitt had characterized minimalism by saying that truth and falsity are terms for going on talking while adding a word or two. Would Jackson et al. then have said ‘Non-cognitivists precisely deny that (e.g.) ethical sentences are used in talking’?

This example may serve to show how easily ideas of correspondence get in through the back door.

10. It seems worth remembering that propounding is a propositional attitude, and that Quinean hostility to propositions is hostility equally to beliefs (say). Because the opposition to certain abstract conceptions of thinkables has typically been directed against things called propositions, we find philosophers whose attitude towards beliefs and statements is one of acceptance, but towards propositions is one of rejection. (See e.g. David 1994, p. 12.) Of course it might be stipulated that the term ‘proposition’ is to mean what is meant by those who use the term illicitly. But short of making such a stipulation, it will be hard to justify an attitude of hostility peculiarly to propositions.

11. Davidson 1967. I use ‘*definition of truth*’ here as Davidson did there; and this allows me to avoid using ‘theory of truth’ ambiguously. (It seems impossible to avoid all possible ambiguity, however. Where a theory of truth [in the only sense of that phrase I use here] purports to give a definition [as the identity theory I defend does not], it purports to give a definition of *truth*; but of course what it purports to give is not a definition of the sort Tarski showed one how to construct, which was a definition of *truth-in-L* for a particular language L.)

One makes no assumptions about Tarski’s own intentions in saying that Tarski in fact showed us a way to construct a definition of truth for L that can be used to do something that a theory of meaning for L has to do. (Etchemendy 1988 has an understanding of Tarski’s purpose which leads to a view of a definition of truth for a language which encourages a deflationary attitude to truth.)

12. I cannot here do more than take for granted a vast body of literature which shows the workability of definitions of truth for languages having natural languages' features. See, e.g. further papers in Davidson 1980. Davidson's idea has been endorsed by many others, of whom, in the present connection, McDowell should be mentioned; see, for example, his 1976.

13. Cp. Davidson 1990, p. 287: 'the concept of truth has essential connections with the concepts of belief and meaning'; and 'what Tarski has done for us is to show in detail how to describe the kind of pattern truth must make'. Davidson himself thinks that the empirical evidence we need in order to identify the pattern must avoid, in the first instance 'states with (as one says) a propositional object'. Davidson, then, would not be happy with the introduction of, 'as one says', propositional objects (i.e. thinkables) at the outset. This explains why his objections to Paul Horwich begin at an earlier point than my own do. For his part, Davidson has a theory of verbal interpretation to elaborate: see 1990. To question the need for this would take me too far afield. But I can try to state Davidson's view in my own terms: such a theory of verbal interpretation has to be understood from the standpoint of someone contemplating an interpretive account in order that such contemplation should ensure that a philosophically adequate conception of truth is elicited.

14. Here I am thinking of, for example, the theory which seems to be endorsed in Putnam's 1981, which says that truth is an idealization of rational acceptability (see chap. 11 in this volume). In later writings (e.g. 1990), Putnam asks us to read the remarks he makes in supporting his Internal Realism as meant only to convey a picture, rather than as a theory of truth.

15. The remarks of this section are intended to go further than those of McDowell (reported in §I.1)—further towards showing that it is not a difficulty for the identity theory that it circumscribes the world using the notion of a thinkable. Although offered in defence of the claim that an identity theorist has a commonsensically realist conception of facts, they are not offered as a defence of any 'Realism' meriting a capital 'R'. In defending his 'Internal Realism' (see n. 14), Putnam's target was 'Metaphysical Realism', a doctrine which the identity theory is evidently also opposed to.

Of course it is possible to think that a defence even of commonsense realism is required: Michael Dummett has long urged this. In his 1990, Dummett thinks of the 'tacit acquisition of the concept [of truth]' as involving 'a conceptual leap; ... just because this is so, it is open to challenge' (p. 200, in 1993 reprinted version). The leap, Dummett says, is one 'we all [made] at an early stage in our acquisition of our mother tongues': it involves a transition from the 'justifiability condition of an assertion to the truth-condition of the statement asserted' (p. 198). Now Dummett's own understanding of the conceptual leap is shown in his speaking of the notion of justification as 'cruder' and of truth as 'more refined'. But Dummett's opponent may resist any picture of the concept of truth as got from something cruder—as if there were something which might be added to justifiability to get truth, so that the child at some stage had to acquire the added extra. (The identity theorist seems bound to resist this, since she cannot allow *truth*'s applicability to

be separated from *thinkability*.) Against Dummett, it may be said that the child who comes to belong to a community of speakers (a systematic account of whose uses of sentences deploys the concept of truth) is drawn into practices in which the concept already has a place. Evidently in saying this, one still does not supply the defence which Dummett seeks. But perhaps it helps to make it clear that one can reject Dummett's story about the acquisition of the concept of truth while acknowledging that truth is indeed in an obvious sense more demanding than justifiability.

16. The reader of David 1994, for example, is invited to accept a correspondence theory of truth on the basis of a demonstration of the untenability of disquotationalism. The dilemmas are sometimes well concealed. For instance, "robustness" may be taken to accrue to truth, or "factualism" to a discourse that is "truth-apt" as soon as some assumed tenet of "minimalism" is denied, and then correspondence conceptions are introduced along with talk of robustness or factualism.

17. Note, 1999: For further discussion of why the identity theory should be incompatible with a deflationary attitude toward truth, see part III (pp. 16–22) of the paper from which the present one has been got by editing and extraction: *Proceedings of the Aristotelian Society* 97 (1997): 1–24. A "deflationary attitude" is held both by minimalists about truth like Horwich (1990 and his paper in this volume) and by pragmatists of Rorty's sort (1995, reprinted in this volume).

## References

- Blackburn, S. 1984. *Spreading the Word*. Oxford: Clarendon Press.
- Brandom, R. 1994. *Making It Explicit*. Cambridge: Harvard University Press.
- Candlish, S. 1995. "Resurrecting the Identity Theory of Truth." *Bradley Studies* 1: 116–124.
- David, M. 1994. *Correspondence and Disquotation*. Oxford: Oxford University Press.
- Davidson, D. 1967. "Truth and Meaning." Reprinted in Davidson 1984.
- Davidson, D. 1969. "True to the Facts." Reprinted in Davidson 1984.
- Davidson, D. 1984. *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.
- Davidson, D. 1990. "The Structure and Content of Truth." *Journal of Philosophy* 87: 279–329.
- Davies, M. 1981. *Meaning, Quantification, Necessity*. London: Routledge and Kegan Paul.
- Dodd, J. 1995. "McDowell and Identity Theories of Truth." *Analysis* 55: 160–165.
- Dodd, J. 2000. *An Identity Theory of Truth*. New York: St. Martin's Press.

- Dodd, J., and Hornsby, J. 1992. "The Identity Theory of Truth: Reply to Baldwin." *Mind* 101: 319–322.
- Dummett, M. 1976. "What Is a Theory of Meaning? II." In *Truth and Meaning: Essays in Semantics*, ed. G. Evans and J. McDowell. Oxford: Clarendon Press.
- Dummett, M. 1990. "The Source of the Concept of Truth." In George Boolos, ed., *Meaning and Method: Essays in Honour of Hilary Putnam*. Cambridge: University Press. Page references are to the version reprinted in Michael Dummett, *The Seas of Language*. Oxford: University Press, 1993.
- Etchemendy, J. 1988. "Tarski on Truth and Logical Consequence." *Journal of Symbolic Logic* 53: 51–79.
- Frege, G. 1918. "The Thought." Quotations are taken from the translation by A. M. and Marcelle Quinton, as reprinted in *Philosophical Logic*, ed. P. F. Strawson. Oxford: Oxford University Press, 1967.
- Grover, D. 1992. *A Prosentential Theory of Truth*. Princeton: Princeton University Press.
- Horwich, P. 1990. *Truth*. Oxford: Blackwell.
- Jackson, F., Oppy, G., and Smith, M. 1994. "Minimalism and Truth Aptness." *Mind* 103: 287–302.
- McDowell, J. 1976. "Bivalence and Verificationism." In Gareth Evans and John McDowell, eds., *Truth and Meaning: Essays in Semantics*. Oxford: Clarendon Press.
- McDowell, J. 1994. *Mind and World*. Cambridge: Harvard University Press.
- Neale, S. 1995. "The Philosophical Significance of Gödel's Slingshot." *Mind* 104: 761–825.
- Putnam, H. 1981. *Reason, Truth, and History*. Cambridge: University Press.
- Putnam, H. 1990. "A Defense of Internal Realism." In *Realism with a Human Face*. Cambridge: Harvard University Press.
- Quine, W. V. 1960. *Word and Object*. Cambridge: MIT Press.
- Quine, W. V. 1970. *Philosophy of Logic*. Englewood Cliffs: Prentice-Hall.
- Rorty, R. 1995. "Is Truth a Goal of Enquiry? Davidson vs. Wright." *Philosophical Quarterly* 45: 281–300.
- Rumfitt, I. 1993. "Content and Context: The Paratactic Theory Revisited and Revised." *Mind* 102: 429–454.
- Russell, B. 1940. *An Inquiry into Meaning and Truth*. London: Allen and Unwin.



---

## Truth as Identity and Truth as Correspondence

Marian David

### 1

According to the standard analysis of belief, the mental state of believing involves a relation between a believer and a *proposition*. The latter is an object of a somewhat special sort. If you believe that bats are mammals, then there is something you believe. *What you believe* is said to be the proposition *that bats are mammals*. The proposition is the object of the belief relation and the *content* of your belief state: propositions are “content objects.” Moreover, propositions can be shared. If you and I both believe that bats are birds, then we believe the same thing; we both stand in the belief relation to the proposition that bats are birds. The propositional analysis of belief extends to many other states and acts, including thinking, judging, assuming, asserting, stating, and saying: they are all construed as relating us in various ways to propositions. The analysis comes with the view that propositions are the primary bearers of truth and falsehood. The truth of a belief (thought, statement, etc.) is derived from the truth of the proposition that is its content. If you believe that bats are mammals, then your belief is true iff (if and only if) the proposition that bats are mammals is true. In general, a true belief is a belief that has a true proposition as its content. Propositions are also bearers of broadly logical properties and relations (entailment, inconsistency, necessary truth, etc.), for they are all tied up with truth.

What, then, is it for a proposition to be true? According to the *correspondence* theory of truth, a proposition is true iff it corresponds to a fact and false iff it does not correspond to any fact. The correspondence theory has its competitors. One of them, the *identity theory*, offers a

surprising simplification. It holds that true propositions do not correspond to facts, they *are* facts, and vice versa:

(IT) For every  $x$ ,  $x$  is a true proposition iff  $x$  is a fact.

The label “identity theory” seems apt, for (IT) entails that a true proposition is identical with a fact. Note that (IT) does not actually invoke the concept of *identity*. But this concept is easily incorporated by expanding ‘is a fact’ to ‘is identical with a fact’. The two formulations are best regarded as notational variants of the same view. But the identity theory should not be confused with its weaker sibling: ‘A true proposition is (identical with) a fact.’ This leaves room for lots of facts that are not identical with true propositions; it is entailed by, but does not entail, the identity theory.

A version of the identity theory was advocated early in the twentieth century by G. E. Moore: “Once it is definitely recognized that the proposition is to denote, not a belief or form of words, but an object of belief, it seems plain that a truth differs in no respect from the reality with which it was supposed merely to correspond: e.g., the truth that I exist differs in no respect from the corresponding reality—my existence” (Moore 1901–1902, 21).<sup>1</sup>

At the time at which Moore advocated this view, he also maintained that truth is at bottom indefinable. The same goes for Russell (1904, 75 f.) and Frege (1918, 60). I have formulated (IT) as a basic principle, or “theory,” about truth, rather than as a definition, to make room for this stance. Note however that (IT), unlike its weaker sibling, offers an equivalence; hence, it has at least the right form for an answer to the question ‘What is truth?’ But one may well wonder whether (IT) is aptly referred to as a theory of *truth* rather than facts. Looking at the equivalence, one tends to experience a Gestalt switch. Is this supposed to be a claim about truth, or is it supposed to be a claim about facts? It seems hard to tell. Moore himself inclined towards the latter view, and Frege’s wording suggests an account of facts rather than truth: “What is a fact? A fact is a thought that is true” (1918, 74).

For present purposes, it is not crucial to decide whether (IT) can possibly serve as a definition, or even whether it can be regarded as a theory about truth rather than facts. We can simply take it as a principle

expressing a (mini) theory about truth and facts. The point is that, regardless of what stance one takes on this issue, if (IT) is a true principle about truth (and facts), then the correspondence theory of truth would appear to be mistaken. The two theories seem incompatible. Chisholm saw things differently: "There is no question, then, about the sense in which true propositions may be said to 'correspond with' facts. They correspond with facts in the fullest sense possible, for they are facts" (1977, 88). Chisholm construed identity as a limit case of correspondence. Moore (1901–1902) and Frege (1918), on the other hand, regarded the identity theory and the correspondence theory as competitors. Since they rejected the latter on the grounds that a genuine correspondence between different items is impossible, they evidently assumed that the very notion of correspondence presupposes the nonidentity of the corresponding items. Although it is partly a verbal issue, it seems best to follow Moore and Frege on this point. Correspondence theorists normally hold that truths are not to be identified with facts; they want to say that facts are *truth makers*—that truths are true *because* of the facts, that they are *made true* by facts. Such claims would be pointless if identity were counted as a correspondence relation in the intended sense: it is pointless to say that truths are made true by themselves.

The identity theory will strike many as rather odd. How would one arrive at such a strange view? No doubt, one may have deep metaphysical reasons having to do with the (alleged) impossibility of genuine correspondence between different things. But I think that one of the primary sources for the view lies elsewhere. I think the identity theory is something that emerges quite naturally from how truth talk and fact talk interact with the use of 'that'-clauses—it emerges naturally, that is, if one has embraced the propositional analysis of belief. The propositionalist holds that the 'that'-clause in 'S believes that bats are mammals' refers to the proposition expressed by its embedded sentence; it is a perspicuous name (description, specification) of its referent, for it identifies (specifies) the proposition referred to while referring to it. We can also use ordinary definite descriptions to refer to propositions, e.g., 'Susan's favorite proposition', and if we wanted to, we could baptize propositions with proper names. But these means of referring to propositions are not perspicuous in the way 'that'-clauses are; they do not automatically show



us which proposition is being referred to. Moreover, unlike ordinary descriptions, a 'that'-clause always provides an essential description (specification) of its referent. Assume that Susan's favorite proposition is the proposition that bats are mammals. This proposition may well have failed to be Susan's favorite proposition, but it could not have failed to be the proposition that bats are mammals—'that'-clauses "rigidly designate" the propositions they refer to.

Let us indicate a 'that'-clause with the schematic expression 'that  $p$ ', in which the dummy letter ' $p$ ' can be replaced by any arbitrary declarative sentence. Consider sentences of the form 'It is true that  $p$ '. They do not look like subject-predicate sentences, for the 'that'-clause does not seem to play the role of a subject expression. They look like the results of applying the operator 'it is true that' to a sentence. However, a propositionalist must maintain that the grammatical surface structure of such sentences is misleading. He wants to capture inferences of the form 'she believes that  $p$ ; it is true that  $p$ ; therefore, she believes something that is true' (i.e., for some  $x$ , she believes  $x$  and  $x$  is true—where the objectual variable ' $x$ ' ranges over propositions). To capture such inferences, he must hold that the underlying logical form of 'It is true that  $p$ ' is revealed by recasting it into subject-predicate form: 'That  $p$  is true'. Here the 'that'-clause appears as logical subject referring to a proposition; it has been fused with a predicate, ' $x$  is true', to yield a subject-predicate sentence attributing the property of being true to the proposition referred to by the 'that'-clause. The propositionalist analysis has brought the 'that'-clause into *referential position*. Consider an example: 'That flies are insects is true'. The position occupied by the 'that'-clause could be occupied by a proper name referring to a proposition. Moreover, the position is accessible to quantification, that is, 'That flies are insects is true' implies 'There is some  $x$  such that  $x$  is true' and it is implied by 'For every  $x$ ,  $x$  is true'.

Once the propositionalist treatment of 'that'-clauses is in place, there is a smooth transition from the ordinary use of 'that'-clauses in truth-and-fact talk to the identity theory:

- (a) It is true that  $p$  iff it is a fact that  $p$ .
- (b) That  $p$  is true iff that  $p$  is a fact.
- (IT) For every  $x$ ,  $x$  is a true proposition iff  $x$  is a fact.

The movement starts with schema (a), which records an elementary observation involving truth. Its substitution instances (the results of substituting declarative sentences for the dummy letter '*p*') are obviously and necessarily true. With propositionalist recastings, schema (a) turns into schema (b), which, when generalized, yields the universal generalization expressing the identity theory of truth.

Given the propositional analysis of belief, the movement from (a) to (IT) records a natural progression of thought, leading quite effortlessly into the identity theory of truth. But this derivation is not a deduction. The step from (a) and (b) to (IT) relies on the tacit assumption that all the 'that'-clauses involved in the substitution instances of (b) refer to propositions (and on the assumption that the ones on the left-hand sides always refer to the same propositions as the ones on the right-hand sides). The assumption is a natural one to make for a propositionalist—he may even find himself embracing the more general principle that *every* 'that'-clause refers to a proposition (provided its embedded sentence makes sense). After all, what would become of the propositional analysis of belief if 'that'-clauses were referentially unstable, if they were prone to shift reference from propositions to other things? We can register a conclusion that will come in handy later on. The derivation of the identity theory essentially involves three elements: schema (a), the propositional analysis of belief (which takes us from (a) to (b)), and the assumption that 'that'-clauses have sufficiently stable reference (which takes us from (b) to (IT)). Since schema (a) seems untouchable, it follows that a correspondence theorist who wants to remain within the propositionalist framework has to reject the assumption that 'that'-clauses stably refer to propositions if she is to avoid seeing her theory collapse into the identity theory. More specifically, she has to reject at least the assumption that the 'that'-clauses involved in truth-and-fact talk have stable reference to propositions.

It is instructive to consider how the identity theory differs from a deflationary theory of truth, like the one advocated by Horwich (1998), which focuses on the equivalence schema (E):

(E) The proposition that *p* is true iff *p*.

To put it roughly, Horwich's deflationary theory consists in the infinite collection of propositions that are expressed by the substitution instances

of (E), where a substitution instance of (E) is a sentence that results from substituting for '*p*' a declarative sentence of English or of some possible extension of English. There are two basic differences between the two theories. First, the deflationary theory is not committed to facts. Second, it does not offer a universal generalization about truth; instead, it "offers" an infinite collection of particular propositions about truth. Unlike the identity theory, the deflationary theory is infinite. The theory itself cannot be stated; it can only be circumscribed by reference to schema (E). Moreover, it cannot be turned into a general principle. Unlike schema (b), from which (IT) was derived by generalization, schema (E) does not even yield a well-formed general claim about propositions, for there is nothing referred to on its right-hand side for the variable *x* to range over (it is not of the form '*x* is true iff *x* is ...', i.e., the dummy letter '*p*' on the right-hand side of (E) is not in referential position). We may also note that schema (E) suffers from a peculiar problem of ontological increase. Take any substitution instance of (E): its left-hand side implies an existence claim not implied by its right-hand side, namely a claim of the form 'The proposition that *p* exists'. (IT) has no such problem: moving from either one of its sides to the other leaves the number of objects in the universe constant.

Moore rejected the identity theory a few years after he had proposed it (Moore 1953, 308). His objection is easily stated. Assume that the proposition that *p* is (contingently) true. According to the identity theory, the proposition exists whether it be true or false. But the *fact* that *p* would not have existed if the proposition had been false. Hence, the fact that *p* cannot be identical with the proposition that *p*. The argument is fallacious. With such an argument one could "prove" that husbands are not identical with married men. Say that John is Mary's husband. John would have existed even if he had not been married to Mary. The husband of Mary would then not have existed. It does not follow that Mary's husband is not identical with John. All that follows is that the person who actually is the husband of Mary might not have been the husband of Mary. Similarly, from 'the fact that *p* would not have existed if the proposition that *p* had been false', it does not follow that the fact is not identical with the proposition. All that follows is that the object (the proposition) that actually is the fact that *p* might not have been a fact.

Hence, it might not have been the fact that *p*. In other words, if the proposition had been false, it would not have been correctly describable as ‘the fact that *p*’. The response is essentially due to Cartwright (1987, 76–78). He points out that Moore’s objection assumes that an expression of the form ‘the fact that *p*’ is a rigid designator, designating the proposition that *p* in every world in which it exists. This is question begging, since the identity theorist will hold that an expression of the form ‘the fact that *p*’ is nonrigid, designating the proposition only in those worlds in which it is true.<sup>2</sup>

But the identity theory does have a problem with *falsehood*, and advocates of the theory are often suspiciously brief when it comes to falsehood.<sup>3</sup> One might think, at first, that (IT) already handles falsehood as the negation of truth: for every *x*, *x* is a true proposition iff *x* is a fact, and *x* is a false proposition iff *x* is not a fact. But no, for this would entail that everything there is is a proposition, including G. E. Moore and every part of every proposition. There is the temptation to “go Meinongian” and to say that a false proposition *is* a fact that does *not exist* (and a true one is a fact that does exist). But there are no facts that do not exist; hence, as Plato pointed out (*Theaetetus* 188–189), this version of the identity theory has the consequence that we cannot think what is false. For when you think what is false, then, on this view, what you think is something that does not exist. Hence, it does not exist. Hence, your thought has no content. Hence, you are not thinking anything at all. To handle falsehood in a more reasonable manner, an identity theorist should propose one of the following:

- (ITa) For every *x*: *x* is a true proposition iff *x* is a fact, and *x* is a false proposition iff *x* is a proposition that is not a fact.
- (ITb) For every *x*: *x* is a true proposition iff *x* is a proposition that is a fact, and *x* is a false proposition iff *x* is a proposition that is not a fact.

(ITa) makes the smaller addition to (IT) but treats truth and falsehood unevenly. (ITb) treats them evenly but leaves behind the pristine simplicity of (IT). Moreover, it has the consequence that it allows for facts that are not propositions. This is noteworthy, but it seems to go against the spirit of the original. Both (ITa) and (ITb) help remind us of a point that

was already implicit in the original. When you think what is false but might have been true, then what you think is not a fact but, according to the identity theory, it may well have been a fact. It seems that truths and falsehoods should be made from the same stuff. If the contents of true thoughts are facts, then the contents of false thoughts must already be made from the same kind of stuff that facts are made of. Or should we turn this the other way around? When you think what is true but might have been false, then, according to the identity theory, the fact that is the content of your thought might not have been a fact; it might have been a mere content of thought. Does this suggest that facts must be made from the same kind of stuff that mere thought contents are made of?

## 2

On the standard analysis of belief, the contents of our beliefs and thoughts are propositions. When the identity theory of truth is combined with this analysis, the result is rather startling: true propositions *are* facts. Hence, the content of the true thought that *p* is the fact that *p*—the fact itself, not some stand-in or representative of that fact. But isn't this more than a little bizarre? We think of facts as belonging to, or rather constituting, *the world*. The identity theory evokes the picture of the world itself entering the mind—or is the picture rather one of the mind enveloping the world? Or are we being told that the world is constituted by the mind?

It is sometimes said that a correspondence theory of truth opens up a “gap” between our thoughts and reality—a gap that, once opened, turns out to be unbridgeable, which makes it impossible to see how our thoughts could ever match reality. In a similar vein, it is sometimes said that a correspondence theory would make the attainment of knowledge impossible, because the confirmation of a belief would require an impossible comparison between a thought in the mind and a fact of the world. Setting aside the question of how much force such worries actually have against correspondence theories (not much, I would say), the identity theorist might claim that they have no force whatsoever against his theory: if the content of a true thought *is* a fact, the whole issue of matching or comparing thought content with fact can never arise in the

first place. The theory has nice consequences for the metaphysics of mind and knowledge.<sup>4</sup>

But are these nice consequences (if they are really there) not bought at too high a price? Well, what the identity theory really amounts to will depend very much on the underlying view of the nature of propositions and facts. Facts are naturally thought of as composed of worldly objects, properties, and relations. But this must be qualified right away. Facts cannot be “composed” of their constituents in the same sense in which ordinary wholes are composed of their parts. When you have all the parts, then you have the whole. But John, Mary, and the relation of loving are not sufficient to make the fact that John loves Mary. In addition, John must also stand in the relation of loving to Mary. Moreover, the same parts cannot make two different wholes at the same time. But John, Mary, and loving can constitute, at the same time, the fact that John loves Mary and the fact that Mary loves John. So facts are complexes that are not entirely reducible to their constituents: they enjoy a nonmereological mode of composition from worldly objects, properties, and relations.<sup>5</sup>

What about propositions? Chances are that propositions will also be composed in some nonmereological manner. But what are their constituents? Note first that the propositional analysis of belief does not tell us anything about the inner makeup of propositions. It only provides us with relational properties of propositions: they, or at least very many of them, must be possible contents of belief states. One can use points familiar from Frege to argue that propositions must be composed of *concepts*—where a concept should be taken as a way of conceiving of a thing or a property.<sup>6</sup> Concept-propositions help explain how it is possible for *S* to believe that Muhammad Ali is a boxer, even though *S* does not believe, or disbelieves, that Cassius Clay is a boxer: *S* conceives of the same person in different ways, i.e., two different concept-propositions are involved. As far as this type of argument is concerned, concepts could be mental entities or episodes. They could be Descartes and Locke’s immaterial *ideas*, which make their living in our immaterial souls. But if concept-propositions are composed of ideas, then the identity theory is either wildly wrong, or it leads to a radical form of idealism, according to which facts, and hence the world, are composed of soul stuff. If concepts

are Hobbes's material ideas, then they are brain states, and the world is in our head.

However, as Frege repeatedly pointed out, concept-propositions cannot be composed of ideas anyway. For the propositional analysis requires that different persons can have beliefs with the same content at different times. Since ideas are private, they cannot fill the bill. This suggests that concepts should be thought of as *types* of ideas: different persons can have different token ideas of the same type at different times. Concept-propositions would then be *objective* entities, constituted by objective ways of conceiving of things and properties. On this view, which has strong affinities with Frege's own, propositions are abstract entities, not easily localized in space or time. They exist independently of individual thinkers. Hence, the *contents* of our beliefs and thoughts are *external* to our minds, not quite in the same sense in which mountains and bats are external to our minds, but they are surely not internal in the sense in which sensations and token ideas are internal to our minds (strangely, this view of content is often classified as internalism). If one thinks of propositions along these lines, one cannot seriously object to the identity theory of truth on the grounds that it externalizes the contents of our thoughts. However, combined with the identity theory, the view will lead to a peculiar conception of facts and the world. Although it will construe facts as objective, mind-independent entities, it will see them as composed from objective concepts of objects and properties, rather than from the objects and properties themselves. Consequently, the fact that Cassius Clay is a boxer will come out as a different fact than the fact that Muhammad Ali is a boxer—rather peculiar indeed. Concept-propositions are too “fine-grained” to make good facts.

There are widely discussed arguments to the effect that the Fregean approach to content must be mistaken. They can be understood as relying on a principle that ties the notion of content to the notion of truth: *necessarily,  $x$  has the same propositional content as  $y$  only if  $x$  has the same truth value as  $y$ .* The principle codifies the idea (accepted by Fregeans) that, whatever content is, the content of a belief must be something that determines the truth value of the belief. Kripke (1972) points out that the thoughts expressed by ‘Aristotle was born in Stagira’ and ‘The author of the *Metaphysics* was born in Stagira’ might have had different truth values

(the *Metaphysics* might have been written by Theophrastus). Given the above principle, it follows that the two thoughts have different contents. There seems to be no way of conceiving of Aristotle (plausibly available to us) that accounts for Kripke's observation, and this has suggested to many that Aristotle *himself* is a constituent of the thought that Aristotle was born in Stagira. Putnam (1975) offers his Twin Earth thought experiment to make a different point concerning natural-kind terms. Twin Earth is just like Earth in all respects, including the neurophysiological makeup of its inhabitants, except that the liquid that looks and tastes exactly like water is XYZ instead of H<sub>2</sub>O. Since the two substances are qualitatively indistinguishable, an inhabitant of Twin Earth *anno* 1750 will conceive of XYZ in the same way in which his twin on Earth conceives of water. Nevertheless, the beliefs expressed by the Twin Earthling's 'water'-sentences will be made true or false by XYZ, whereas the beliefs expressed by the Earthling's 'water'-sentences will be made true or false by H<sub>2</sub>O. Imagine an Earthling and a tourist from Twin Earth both pointing at Lake Michigan while uttering the words 'This is water': their respective beliefs would have different truth values, which implies that they do not have the same content. Finally, Kaplan (1977) points out that, when uttered by Carol, 'I am hungry' will often have a different truth value than when uttered by Bob, which again implies that the thoughts expressed have different contents.

This view of content seems made to order for the identity theory of truth. It makes content external—in a stronger sense than the Fregean view, for it makes content dependent on the thinker's *spatiotemporal* environment—and it suggests that the contents of our thoughts are composed of objects and properties, which fits our intuitive view of the constituents of facts. Of course, since our thoughts can be false, we have to accept that there are entities that are just like facts, only they are not facts: *that glass is made of water* would be such an entity. Chisholm's formulation of the identity theory (1977, chap. 5) goes well with this view. He said that propositions (next to events) are a species of *states of affairs*: a proposition is true iff it is a state of affairs that obtains and false iff it is a state of affairs that does not obtain. States of affairs that obtain, he said, are facts. States of affairs that do not obtain, we might say, are nonfacts.<sup>7</sup> Of course, being made to order for the identity theory comes



with a price. We have now returned to the view that worldly facts are the contents of our thoughts, when they are true, and the constituents of facts—Mount Everest, Alpha Centauri, Muhammad Ali, fleas, avocados—are the constituents of the contents of our thoughts, be they true or false.

Still, this strongly externalist view of content has been well argued for. So it seems that the identity theory of truth, however bizarre it might have appeared initially, has been vindicated by relatively firm results in the theory of content. Moreover, some of the air of paradox surrounding the identity theory can be dispelled by pointing out that it derives from the use of the term ‘content’. The term was inherited from the propositional analysis of belief, which refers to propositions as the contents of belief states, as if a belief state were some sort of container. But surely this is a bit metaphorical. The term is just a label that has been associated with ‘that’-clauses and with the phrase ‘what *S* believes’. Why not change metaphor and say that propositions are *objects* of belief? We would normally refer to Mary as the object of John’s love. So if *S*’s state of believing that *p* is supposed to be a *relation* to a proposition, should we not refer to that proposition as the object of *S*’s belief?<sup>8</sup> Note that we do not find it at all bizarre to refer to Mount Everest as the object of a belief. Thinking of propositions and facts as objects of belief might help lower resistance to the identity theory (although it seems questionable that there is a sense of ‘object’ in which Mount Everest and the fact that Mount Everest is a mountain are both objects of the belief that Mount Everest is a mountain). But there is a more serious problem for the identity theory. The strongly externalist view of content does not fully vindicate the identity theory, for the arguments supporting strong externalism apply only to proper names, demonstratives, and natural-kind terms. Only thought contents completely expressible with such terms are strongly external. Artifact concepts and other functional concepts are not “Twin-Earthable”: a Twin Earth *beverage* is a beverage, even though it contains XYZ. Moreover, the arguments do not even work for *complex* concepts of natural kinds: if hydrogen and oxygen exist on Twin Earth, then a Twin Earthling can think of hydrogen and oxygen; he can then go on to form the concept of H<sub>2</sub>O in the absence of water.<sup>9</sup> So, strong externalism applies only to a subset of our thought contents, and hence only to a small subset of the propositions we can think. But the identity theory

says that *every* true proposition is a fact (and every false proposition is constituted just like a fact, only it is not a fact). Strong externalism offers only a very partial vindication of the identity theory of truth.

Let us take stock. Given the propositional analysis of belief, propositions cannot be mental entities. Even if they could be, combining mentalism about the nature of propositions with the identity theory of truth would lead into radical idealism. The Fregean account of the nature of propositions is a form of externalism about content, according to which propositions are complexes of objective ways of conceiving of objects and properties. Combining this with the identity theory leads to a distinctly peculiar view of the nature of facts. The strongly externalist Kripke-Putnam-Kaplan-inspired view of content, combined with the identity theory, leads to a reasonable view about the nature of *some* facts, but strong externalism does not cover all our thought contents—not by a long shot. It is tempting to opt for a mixed view of content: let strong externalism about content reign as far as it can, and let some version of Fregean concept-propositions cover the remaining ground. This will require allowing for mixed propositions, constituted partly by worldly objects and properties and partly by objective ways of conceiving of objects and properties. This looks promising as far as the theory of content is concerned, but as far as the identity theory of truth is concerned, adopting such a mixed view will not help much. The identity theory would still require a huge stock of peculiar facts.

### 3

So let us take a look at the correspondence theory. (I do not mean to imply that failure of the identity theory should count as evidence for the correspondence theory. I am merely restricting my attention to these two approaches.) In particular, let us look at a correspondence theory that would be a direct competitor for the identity theory, i.e., one that talks about correspondence between facts and propositions, where the latter are understood in a sense faithful to the propositional analysis of belief. Let me remind you of the conclusion I drew in the first section when commenting on the “derivation” of the identity theory. The correspondence theorist who wants to be faithful to propositionalism without

seeing her view collapse into the identity theory has to reject an assumption that comes very naturally, once the propositional analysis is in place. The assumption she has to reject is this: the ‘that’-clauses involved in truth-and-fact talk have stable reference to propositions. Rejecting this will involve rejecting the idea that the ‘that’-clauses in substitution instances of schemata like (a) (‘It is true that  $p$  iff it is a fact that  $p$ ’) and (b) (‘That  $p$  is true iff that  $p$  is a fact’) always refer to the same thing.

Consider the following objection to the identity theory. The theory commits one to the view that facts are true: every fact is a true proposition; every true proposition is true; hence, every fact is true, which is absurd. The identity theorist will want to take this in stride. She will want to say that ‘Facts are true’ is literally true; it merely sounds odd because it amounts to the redundant claim that true propositions are true. Objections like the above can always be “met” in this manner, i.e., by maintaining that what the objection declares to be a *reductio* merely sounds odd but is actually true. And this type of response is not necessarily without merit: philosophers do find themselves at times saying odd things for fairly good reasons. But the present situation is a bit special. For the derivation of the identity theory, which stands in the way of the correspondence theory, relies itself on a linguistic claim about truth talk and fact talk. Hence, it does seem quite appropriate to point out that truth talk and fact talk do not go together as smoothly as the identity theorist might like. Austin (1961) offers a number of cases, e.g., we can say ‘It is true to say that flies are insects’ and also ‘To say that flies are insects is true’, but we cannot say ‘It is a fact to say that flies are insects’ or ‘To say that flies are insects is a fact’. A more common observation is that facts can apparently be causes: ‘The panic was caused by the fact that the theater was on fire’. But propositions and their near relatives (statements, judgments, etc.) are not causally efficacious. Also, facts can be discovered and disclosed, and explaining the fact that the Dodo is extinct involves activities rather different from explaining the proposition or statement that the Dodo is extinct.

There is, then, ample evidence for thinking that truth talk and fact talk do not go smoothly hand in hand. But pointing this out is not enough. Something needs to be said about how truth talk and fact talk relate; in particular, the role of ‘that’-clauses needs to be clarified. Let us begin

with singular terms of the form 'the proposition that *p*' and 'the fact that *p*'. At first glance, they look like definite descriptions. But they are not plausibly interpreted as ordinary definite descriptions. The phrase 'the proposition that *p*' does not dissolve in Russellian manner into: 'there is one and only one proposition *x* such that *x* is a proposition that *p* and *x* is ...'. Let us call such descriptions quasi descriptions. How do they work? Compare them with 'the planet Jupiter' and 'the god Jupiter'. These do not work like ordinary definite descriptions either. The ordinary description 'the planet beyond Jupiter' refers to a thing other than Jupiter by relating it to what the embedded name 'Jupiter' refers to. But 'the planet Jupiter' refers to the same thing as the embedded name 'Jupiter'. Such quasi descriptions are useful because 'Jupiter' is an ambiguous name, and they allow us to disambiguate the embedded name; 'the planet ...' and 'the god ...' tell us how to take the name 'Jupiter'. Moreover, unlike ordinary descriptions, they can be turned into subject-predicate sentences without much ado: 'the planet Jupiter' turns directly into 'Jupiter is a planet', in which the predicate serves to disambiguate the name retroactively.

Now, the quasi descriptions 'the proposition that *p*' and 'the fact that *p*' work exactly like that. That is, the 'that'-clause, 'that *p*', is *ambiguous*, it refers to one type of thing when preceded by 'the proposition ...', namely to a proposition, and to another type of thing when preceded by 'the fact ...', namely to a fact, provided there is an appropriate one. The quasi description 'the proposition that *p*' refers to the proposition expressed by the sentence embedded in the 'that'-clause. What does the quasi description 'the fact that *p*' refer to? To a fact. To which fact? It refers to whatever fact functions as truth maker for the proposition expressed by the sentence embedded in the 'that'-clause. If that proposition has no truth maker, then the expression does not refer. Since 'the fact that *p*' is a quasi description, the embedded ambiguous 'that'-clause refers to the same thing, if any, as the whole quasi description. And since quasi descriptions can be turned into subject-predicate sentences in which the predicate has to do the disambiguating work retroactively, our two quasi descriptions are readily transformed into 'That *p* is a proposition' and 'That *p* is a fact'. The 'that'-clauses refer to different things. And since the 'that'-clause in 'That *p* is a fact' refers to a fact if and only if its

embedded sentence expresses a true proposition, 'That  $p$  is a fact' is equivalent to 'That  $p$  is a true proposition'. This accounts for schema (b)—that  $p$  is true iff that  $p$  is a fact. It also accounts for schema (a), for, according to propositionalism, they are notational variants of each other.

The correspondence theorist can reject the assumption that 'that'-clauses have stable reference in truth-and-fact talk. He can give a workable account of why and how they switch referents from propositions to facts. But to get a plausible correspondence theory with reasonably worldly facts, the correspondence theorist needs to do more. Take a simple correspondence account: a proposition is true iff it corresponds to a fact and false iff it does not correspond to a fact. We have seen in the previous section that propositions will likely be a varied lot. There will be propositions composed of objects and properties, propositions composed of concepts, and mixed propositions composed partly of objects or properties and partly of concepts. Moreover, there will be logically complex propositions whose immediate constituents are simpler propositions. A reasonable correspondence theory should make each true proposition correspond to a worldly fact composed of objects, properties, and relations.

The basic outlines of how this should go have been provided by advocates of *atomistic* correspondence theories of truth, first proposed by Wittgenstein (1921) and Russell (1918), and later modified and developed by David Armstrong (1997) and others. The first step of this approach is to uphold the *truth-maker principle*: for every truth there must be something in the world that makes it true: every true truth bearer must have a truth maker.<sup>10</sup> The next step is to reject the tempting idea that there is a one-one correspondence between truth bearers and truth makers. Instead, the relation is many-many: one proposition can be made true by different truth makers, and different propositions can be made true by one truth maker. Disjunctive propositions make for an instructive example: a disjunctive proposition is true iff one or both of its disjuncts is true. Consequently, there is no need for a disjunctive fact to make the disjunction true, and there is no need to find a worldly counterpart for the logical concept expressed by 'or': "My fundamental idea is that the 'logical constants' are not representatives; that there can be no representatives of the *logic* of facts" (Wittgenstein 1921, 4.0312). The hope is

that all logically complex propositions can be handled in some such manner, so that there is no need for any facts but atomic facts; they are the sole truth makers, though conjunctive facts are usually permitted because they are mere aggregates of atomic facts. So far, this strategy would still leave the correspondence theorist with a large class of elementary propositions that are completely or partly composed of concepts (elementary propositions are those that do not have any propositions as constituents). So the atomist will go on to reject the tempting idea that there is a one-one correspondence between predicative concepts and genuine universals. Instead, one concept may correspond to one, many, or no universals, and one universal may correspond to one, many, or no concepts. So, for example, the proposition that Peter is playing a *game* can be made true by any one of a variety of facts involving particular games. And the fact that John loves Mary makes true the proposition that John loves Mary, and the proposition that John loves someone, and many more. While Russell and Wittgenstein may have held that the nature of the constituents of atomic facts is to be determined on the basis of a priori considerations, Armstrong advocates a posteriori scientific atomism. On his view, atomic facts are composed of particulars and simple universals (properties and relations). The latter are objective features of the world that ground the objective resemblances between particulars and explain their causal powers. Accordingly, what particulars and universals there are will have to be decided on the basis of total science. There are, of course, difficulties with working out the details, and concessions may have to be made—e.g., true negative propositions pose problems for an atomist who wants to avoid commitment to negative facts and wants to give an acceptable account of falsehood.

Facts, even worldly atomic facts, are often spurned. Because they are specified by using 'that'-clauses, they are said to be "projected" from true propositions for the sake of correspondence. Surely this complaint is rooted in the (tacit) assumption that 'that'-clauses have stable reference to propositions and in the ease with which the mind moves from this assumption towards the identity theory. A reasonable correspondence-to-fact theory will not get this movement going. It has to be conceded that quasi descriptions of the form 'the fact that *p*' will often express mere concepts that correspond or refer rather messily to any number of atomic

facts. However, this does not hold whenever 'that  $p$ ' refers to an atomic fact. I have not made a case for facts here, because the identity theory already came equipped with them. A friend of facts will argue that a mere particular  $a$  is not sufficiently *articulated* to serve as an adequate truth maker. If  $a$  were the sole truth maker of the proposition expressed by ' $a$  is  $F$ ', then ' $a$  is not  $F$ ' would have to be true too. So the truth maker for ' $a$  is  $F$ ' needs at least to involve  $a$  and  $F_{\text{ness}}$ . But since  $F_{\text{ness}}$  is a universal, it could be instantiated in another object  $b$ . Hence the mere existence of  $a$  and  $F_{\text{ness}}$  is not sufficient for making true the claim ' $a$  is  $F$ ':  $a$  and  $F_{\text{ness}}$  need to be tied together in the fact that  $a$  is  $F$ . In addition, I would maintain that some facts are observable: I can *see* that the cat is eating. However, the friends of facts should make a concession. Essential predications, like the proposition that Fido is a dog, are not in need of facts as truth makers. Fido is sufficient, for wherever he goes, the universal being a dog must necessarily follow. Maybe correspondence to *facts* should be restricted to contingent truths.

It is sometimes said that correspondence between propositions and facts cannot be a full-fledged relation, and there is something to that. Note first that on the present view, correspondence must be a generic concept that refers to a number of different relations. After all, if content externalism is correct, then there will be true propositions for which correspondence shrivels to identity. But since there are vast numbers of concept-propositions, this will not be the typical case. Still, the correspondence relation is not quite full-fledged, even with respect to concept-propositions. According to the standard understanding of the propositional analysis of belief, the correspondence relation will be an internal relation, for the existence of its relata (say, the proposition that John loves someone and the fact that John loves Mary) entails that the relation holds between them (see Armstrong 1997, 129). Moreover, unlike resemblance, which is also an internal relation, correspondence between propositions and facts appears to be a relation that is wholly founded in the essential nature of its relata (the proposition could not have failed to be the proposition that John loves someone, and the fact could not have failed to be the fact that John loves Marry, without going out of existence, that is). So in this sense, correspondence is less than full-

fledged. On the other hand, when it relates concept propositions to facts, it relates different items, so it is a more substantial relation than identity.

Finally, I need to dispel an illusion I created for the sake of simplicity. I said in section 1 that the identity theory is not a version of the correspondence theory. This requires a qualification. We can distinguish at least four broad categories of truth bearers: sentences (and the like), statements, beliefs, and propositions.<sup>11</sup> Due to this diversity of truth bearers, it is at best misleading to talk of *the* correspondence theory of truth. There really are four groups of possible correspondence theories—one for each type of truth bearer. It can be consistent to combine a correspondence approach for truth bearers of one type with a different approach for truth bearers of another type. The identity theory is a theory of proposition truth. Although it is not a correspondence theory of proposition truth, it does not follow that it cannot be *part of* a correspondence theory of truth for other truth bearers. A composite theory that combines the two approaches will count as a correspondence theory overall.

A rough account of *sentence truth* will serve to illustrate: (i) a sentence is true iff it represents a true proposition, and (ii) a true proposition is a fact (the identity theory). Although no genuine relation occurs in the second part of this account, there is a full-fledged relation mentioned in the first part: the semantic relation of *representation*—an external relation that holds contingently between its relata. Admittedly, the resulting theory is, as it were, “more compressed” than it would be if (ii) offered a correspondence theory of proposition truth. But it is as a correspondence theory nevertheless, for it is (the beginning of) an explanation of what it is for a sentence to correspond to a fact. There is sufficient “play” in the relation of representation for the composite theory to count as a correspondence theory of sentence truth. The situation is different in the case of *belief truth*. Following the model provided above, a “compressed” correspondence account for belief truth would take the form: (i\*) a belief is true iff it bears a relation *R* to a true proposition, and (ii) a true proposition is a fact (the identity theory). But the noun ‘belief’ harbors an “act-object” ambiguity: at times it is used to refer to the state of believing something; at other times it is used to refer to what is believed. On a



traditional understanding of the propositional analysis of belief, belief *states* are, strictly speaking, not truth bearers at all. We say ‘What she believes is true’; we do not say ‘Her believing it is true’. The form ‘Her belief that *p* is true’ is to be construed along the following lines: ‘The proposition that *p* is true and is believed by her’. So this view denies that there is an intelligible sense in which beliefs (thoughts, judgments, etc.) are truth bearers in addition to propositions.<sup>12</sup> It follows, then, that (i\*) is nonsense if ‘belief’ refers to a belief *state*. And if ‘belief’ refers to a proposition, (i\*) is already covered by (ii): there is no “play” for relation *R*, no room for any form of correspondence. So the identity theory of truth has no use for correspondence when it looks at beliefs. But when it looks at sentences, it is in need of a little help from a friend.

## Notes

1. This commits Moore only to (IT)’s weaker sibling, but the context makes clear that he intended (IT). Other advocates include (Russell 1904, 74–76), Meinong (1910, chap. 3), Frege (1918, 74), Ducasse (1940), and Chisholm (1977, chap. 5). Moore soon discarded the theory. Russell flirted with it only briefly. Frege mentioned it once (he called propositions ‘thoughts’)—but the identity theory seems inconsistent with his overall views on semantics, and the context of his remark leaves open whether he was entirely serious about it. The identity theory has recently received some renewed attention. Candlish (1989), who introduces the label ‘identity theory’, and Baldwin (1991) discuss the version of Bradley, who seems to have favored more “holistic” formulations along the lines of ‘Truth is identical with reality’ (see Candlish 1989, 338). Bradley, it seems, was not given to chopping up *truth* and *fact* into truths and facts. Baldwin offers a nice quote from Hegel: “Truth in the deeper sense consists in the identity between objectivity and the notion” (1830, §213). A version of the identity theory shows up in McDowell (1994, 27). It is the starting point for an extended discussion by Hornsby (1997 [see also chap. 28—Ed.]), but it is unclear whether Hornsby’s official formulation is the weaker sibling or (IT): “The identity theory is encapsulated in the statement that true thinkables are the same as facts” (1997, 2).

2. Note that this does *not* conflict with the idea that ‘that *p*’ in ‘the fact that *p*’ rigidly designates the proposition that *p*.

3. McDowell says, “When one thinks truly, what one thinks *is* what is the case. So ... there is no gap between thought, as such, and the world” (1994, 27). Although he concedes right away that thought can be “distanced from the world by being false,” the no-gap conclusion relies on an inference from what holds only for *true* thought as such to thought as such.

4. This holds true only for the *metaphysics* of knowledge. The identity theory does not imply that we get knowledge somehow for free. Say, *S* believes that *p*,

and that *p* is a fact. It does not even begin to follow that *S* knows that *p*. To think otherwise is to confuse knowledge with true belief.

5. A Platonist about universals (properties and relations) will hold that the relation of loving exists even if no one ever loves anyone. An Aristotelian will hold that the relation exists only if someone loves someone at some time. Facts seem more worldly when composed of objects and Aristotelian universals.

6. See Frege 1892. Frege would have talked of “modes of presentation” and the “senses” of words instead; he used ‘Begriff’ in a different and somewhat strange way.

7. Confusingly, Chisholm construed states of affairs and propositions as Fregean concept entities.

8. See the passage from Moore (1901–1902, 21), quoted in the beginning.

9. This was pointed out by Colin McGinn (see his impressive discussion of externalism in 1989, chap. 1).

10. See Armstrong 1997, chap. 8. In what follows I will often use Armstrong’s terminology. However, where he talks of *states of affairs*, I will talk of *facts*, for in the previous section I briefly used ‘state of affairs’ to refer to a genus of strongly externalist propositions.

11. Propositions are relatively theoretical entities. Consequently, their status as truth bearers differs somewhat from the status of the other items on the list: it is not a *datum* for a theory of truth that there are propositions whose truth or falsehood has to be accounted for.

12. The view does not deny that *sentences* are truth bearers in addition to propositions; however, it will aim to explain sentence truth in terms of representation plus proposition truth.

## References

Armstrong, D. M. 1997. *A World of States of Affairs*. Cambridge: Cambridge Univ. Press.

Austin, J. L. 1961. “Unfair to Facts.” In his *Philosophical Papers*, 3d ed., edited by J. O. Urmson and G. J. Warnock, pp. 154–174. Oxford: Oxford University Press, 1979.

Baldwin, T. 1991. “The Identity Theory of Truth.” *Mind* 100: 35–52.

Candlish, S. 1989. “The Truth about F. H. Bradley.” *Mind* 98: 331–348.

Cartwright, R. L. 1987. “A Neglected Theory of Truth.” In his *Philosophical Essays*, pp. 71–93. Cambridge: MIT Press.

Chisholm, R. M. 1977. *Theory of Knowledge*. 2nd ed. Englewood Cliffs, N.J.: Prentice-Hall.

Ducasse, C. J. 1940. “Propositions, Opinions, Sentences, and Facts.” *Journal of Philosophy* 37: 701–711.

- Frege, G. 1892. "On Sense and Meaning." In his *Collected Papers on Mathematics, Logic, and Philosophy*, edited by B. McGuinness. Oxford: Basil Blackwell, 1984.
- Frege, G. 1918. "Thoughts." In his *Collected Papers on Mathematics, Logic, and Philosophy*, edited by B. McGuinness. Oxford: Basil Blackwell 1984. Page references are to the original pagination.
- Hegel, G. W. F. 1830. *The Science of Logic*. In *Hegel's Logic*, translated by W. Wallace. Oxford: Clarendon Press, 1975.
- Hornsby, J. 1997. "Truth: The Identity Theory." *Proceedings of the Aristotelian Society* 97: 1–24.
- Horwich, P. 1998. *Truth*. 2nd ed. Oxford: Oxford Univ. Press.
- Kaplan, D. 1976. *Demonstratives*. In J. Almog, J. Perry, and H. Wettstein, eds., *Themes from Kaplan*. New York: Oxford Univ. Press, 1989.
- Kripke, S. 1972. *Naming and Necessity*. Cambridge: Harvard Univ. Press 1980.
- McDowell, J. 1994. *Mind and World*. Cambridge: Harvard Univ. Press.
- McGinn, C. 1989. *Mental Content*. Oxford: Basil Blackwell.
- Meinong, A. 1910. *Über Annahmen*. 2nd ed. In his *Gesamtausgabe*, vol. 4, edited by R. Haller and R. Kindinger. Graz: Akademische Druck- u. Verlagsanstalt, 1977. 1st ed., 1902.
- Moore, G. E. 1901–1902. "Truth and Falsity." In *Dictionary of Philosophy and Psychology*. Reprinted in his *Selected Writings*, edited by T. Baldwin, pp. 20–22. London: Routledge, 1993.
- Moore, G. E. 1953. *Some Main Problems of Philosophy*. London: George Allen and Unwin.
- Plato. 1997. *Complete Works*. Edited by J. M. Cooper. Indianapolis: Hackett.
- Putnam, H. 1975. "The Meaning of 'Meaning'." In *Mind, Language, and Reality*, vol. 2 of *Philosophical Papers*, pp. 215–271. Cambridge: Cambridge Univ. Press.
- Russell, B. 1904. "Meinong's Theory of Complexes and Assumptions." *Mind* 13. Reprinted in his *Essays in Analysis*, edited by D. Lackey, pp. 21–76. New York: Braziller, 1973.
- Russell, B. 1918. "The Philosophy of Logical Atomism." In his *Logic and Knowledge: Essays, 1901–1950*, edited by R. C. Marsh. London: George Allen and Unwin, 1956.
- Wittgenstein, L. 1921. *Tractatus Logico-philosophicus*. Trans. by D. F. Pears and B. F. McGuinness. London: Routledge and Kegan Paul 1961.

---

## The Face of Cognition

Hilary Putnam

### Dummettian antirealism

Michael Dummett sees the problem of realism as having to do with the “recognition transcendence” of truth. Either truth is simply the state of being verified, or it transcends what the speaker can verify, he argues, and if it transcends what the speaker can verify, it is not a property whose presence the speaker can “recognize.” And if truth is a property whose presence (in some cases, at least) the speaker cannot recognize, then the speaker’s alleged “grasp” of the notion of truth becomes a mystery. In effect, Dummett is telling us, if truth is not verifiable, then, short of postulating magical powers of mind, we shall not be able to explain how we understand the notion. The rejection of magical powers of mind requires the acceptance of a very radical form of verificationism, according to Dummett’s<sup>1</sup> line of thinking—one so radical that it requires us to revise a number of the laws of classical logic, beginning with the principle of bivalence.

There is a rejoinder to Dummett’s argument which he himself anticipated from the beginning, and which he discusses at length in *The Logical Basis of Metaphysics*. That rejoinder, which in essence goes back to Alfred Tarski’s<sup>2</sup> celebrated essay on the concept of truth, runs as follows.

What is your problem? Take any sentence you like—take a sentence whose truth value we may not be able to find out, if you please. For example, take the sentence:

(1) Lizzie Borden killed her parents with an axe.

Even if the truth of this sentence is ‘recognition-transcendent’, surely you understand what it *means* to say that (1) is true. For you understand (1) itself, and the chief logical principle governing the use of the word ‘true’ is:

[Tarski's Convention T:] If *S* is the name of any sentence, and we write that sentence<sup>3</sup> in the blank in:

(2) *S* is true if and only if \_\_\_\_ then the resulting sentence will be true.

[Less formally: a sentence that says of another sentence *S* that *S* is true is equivalent to *S* itself. Tarski's famous example was:

'Snow is white' is true if and only if snow is white.]

In short, you understand 'Lizzie Borden killed her parents with an axe' and you know that

'Lizzie Borden killed her parents with an axe' is true if and only if Lizzie Borden killed her parents with an axe.

So you do understand what it means to say that (1) is true; *it means that Lizzie Borden killed her parents with an axe.*

I want also to note the fact that some philosophers who offer this account of how we understand sentences of the form '*S* is true'—but not Tarski himself<sup>4</sup>—add the claim that truth is not a "substantive property." These philosophers—I shall refer to them as *deflationists*, in order to distinguish their position from Tarski's own (unmodified) position—claim that the predicate 'is true' is just "a logical device."<sup>5</sup> I shall say something about this "deflationist" position shortly.

Dummett's reply to the (unmodified) "Tarskian" argument takes us to the heart of his philosophical concerns, however. "Granted that I understand sentence (1), and other sentences with an unknown truth value, for example, undecided conjectures in mathematics," he answers [in effect—I am formulating his reply in my own words], "the philosophical problem is *to give an account of what that understanding consists in.*" In short, if you appeal to an *unexplicated* notion of "understanding a sentence," then you are simply ducking all the philosophical problems.

According to Dummett, my understanding of the sentence (1) [that is, of any sentence] consists in my ability to *recognize whether* (1) *is verified*. In other words, *if* (1) *should* be verified (by data that I myself perceive), then I *would* be able to tell that it was; and the ability or system of abilities that enables me to do this *constitutes* my understanding of (1). Similarly, I possess the ability to recognize proofs in mathematics, and this allows me to say that, if I were given a proof of the conjecture that there are infinitely many twin primes (pairs of primes such that one is obtained by adding two to or subtracting two from the other), I could

recognize that it *was* a proof. And that is how I can say that I *understand* the twin-prime conjecture.

Dummett, of course, would concede the “Tarskian” points that he also understands the statement that (1) is true and the statement that the twin-prime conjecture is true, and that he knows that the statement that (1) is true is equivalent to (1) itself, etc. “But notice,” he will point out [my words again!], “If my account is right, a speaker’s understanding of the statement that (1) is true involves the speaker’s understanding what it is for (1) to be verified—and this property, being verified, is a property that (1) and its negation may *both* lack; it does not require the speaker to know anything about a property—call it ‘classical truth’—that must be possessed either by (1) or else by (1)’s negation, independently of whether anyone can tell which one possesses it, as is postulated by classical logic.” In short, if Dummett’s verificationist account of *what constitutes understanding* is right, then either truth is a useless metaphysical abstraction, or else there is nothing to the claim that *truth is a bivalent property*, the claim that characterizes “two-valued” logic. (It is thus that Dummett is led to the radical claim that a sound philosophy of language requires the revision of classical logic itself.)

I want now to consider the response of the “deflationist philosophers” I mentioned a few moments ago. These philosophers agree with Dummett in thinking of our understanding of our sentences as *consisting in* our knowledge of the conditions under which they are verified, although they reject Dummett’s notion of “conclusive verification,” replacing that notion with a notion of degrees of verification.<sup>6</sup> They also reject Dummett’s claim that we must not think of truth as a bivalent property, although they do agree that it is not a “substantive property” about which some metaphysical story needs to be told; rather they claim that rejecting that metaphysical picture of what truth is does not require us to give up the “law of the excluded middle,” “ $p \vee \neg p$ .” As just mentioned, the deflationists even allow us to assert bivalence:

(3) Either  $p$  is true or the negation of  $p$  is true.

where  $p$  is any declarative sentence,<sup>7</sup> but they interpret the assertion of (3) as a mere linguistic practice, free of commitment to the existence of a property ‘truth’ that is determinately possessed either by the sentence or

else by the negation of the sentence. For example, if we put sentence (1) for  $p$ , what (3) means, they say, is

- (4) Either Lizzie Borden killed her parents with an axe or Lizzie Borden did not kill her parents with an axe.

—and (4), it will be noted, does not contain the word ‘true’.

But why should we accept (4)? Deflationists give different answers. Rudolf Carnap and Ayer said that the acceptance of sentences of the form ‘ $p$  or not- $p$ ’ is a linguistic convention; Quine, rejecting that answer, says simply that such sentences are “obvious” (sometimes he says “central” to our reasoning). But does not the “obviousness” of (4) depend on our belief that there is a fact of the matter as to whether Lizzie Borden did or did not administer the famous “forty whacks”? And if uttering a sentence (whether or not I also employ the “logical device” of saying that the sentence ‘is true’) is just following a community-wide practice of assigning it a degree of assertability “as a function of observable circumstances,” how do we so much as make sense of the idea of a fact of the matter as to the rightness of statements that are *neither* confirmed nor disconfirmed by those observable circumstances?

If we structure the debate in the way in which both Dummett and the deflationists do, then we are left with a forced choice between (a) either Dummettian antirealism or deflationism about truth, or (b) a retreat to metaphysical realism. Both Dummett’s “global antirealist”<sup>8</sup> and the deflationist advertise their accounts as rescuing us from metaphysical realism. But, surely, one of the sources of the continuing appeal of metaphysical realism in contemporary philosophy is a dissatisfaction with the only apparent alternatives. The metaphysical realist will want to reply to the deflationist (and the antirealist) as follows.

“Realism requires us to say that either (1) or the negation of (1) is true. If a philosopher advises us to retain ‘Either (1) is true or the negation of (1) is true’ as something we are permitted to say while reinterpreting *what we are doing when we say it* in such a way as to deprive us of what we ordinarily mean (when we say of a sentence that it is true), then he is disguising the radically revisionary character of his theory through a terminological slight-of-hand. That is what the deflationist, in effect, does. He allows us to hold on to the thought that ‘Either (1) is true or the

negation of (1) is true' only in the attenuated sense that he advises us to follow a policy of assigning all grammatical sentences of the syntactic shape ' $p \vee \neg p$ ' the degree of assertability (the 'level of confidence', in Paul Horwich's phrase) 1. This attenuated sense in which the deflationist continues to permit us to speak of a sentence's being true or false fails to capture what is significant about true sentences (as opposed to false ones): true sentences possess a substantive property that false sentences lack—namely, the property of corresponding to a reality. Deflationism is thus unable, for example, to acknowledge the reality of past events (as things that truly happened), even though it retains the old form of words ('It happened or it did not happen') as a *mere* form of words. Deflationism, in effect, follows the lead of logical positivism<sup>9</sup> in refusing to think of our sentences as subject to serious terms of normative appraisal, of appraisal in terms of the possession or absence of a substantive property of rightness that is different from verifiability. On the deflationist account, when one asserts the whole sentence '(1) is true or the negation of (1) is true' one is not saying that one of the *disjuncts* possesses the relevant sort of substantive rightness. The deflationist is unable to do justice to the sense in which one of the disjuncts of this sentence possesses the same sort of substantive rightness as does (if you are presently reading this essay) the sentence 'You are right now reading these words in front of you'. The deflationist (by regarding degree of assertability, but not truth, as a property that is more than just a logical device<sup>10</sup>) is therefore unable to capture the sense in which certain statements about the past (namely, the true ones) are fully as *right* as statements about the present. Dummett perceives the situation more clearly than the deflationists in that he at least recognizes—indeed emphasizes—that his account of understanding commits him to antirealism about the past (and not only about the past). Neither Dummett nor the deflationist, however, can accommodate the ordinary sense in which certain statements about the past are substantively true."

What is the difference between the realism of the metaphysical realist (whose response to deflationism I just sketched) and the common-sense realism that I wish to attribute to Wittgenstein? In a different context (in response to a Platonist about rule-following), Wittgenstein writes,



Really the only thing wrong with what you say is the expression “in a queer way.” The rest is all right; and the sentence only seems queer when one imagines a different language-game for it from the one in which we actually use it.<sup>11</sup>

Wittgenstein would, I believe, reply to the metaphysical realist’s response to the deflationist (which I have sketched above) by saying, “Really the only thing wrong with what you say is the expression ‘substantive property’ (and related uses of ‘substantive’, as in ‘substantive sort of rightness’ and ‘substantively true’).” Thus, from Wittgenstein’s point of view, most of the words that the metaphysical realist finds himself moved to say (in response to the deflationist) are perfectly all right. But the metaphysical realist makes these words seem fated to say something *queer* by calling upon them to bear an explanatory burden—to bear metaphysical weight—in accounting for the relation between thought and reality. The metaphysical realist feels that the deflationist has drained our ordinary ways of speaking and acting of their substance, and so he seeks to reinfuse them somehow with substance. It is to this end that he ineffectually invokes the notion of a ‘substantive property’. The metaphysical realist (in trying to do justice, for example, to our ordinary realism about the past) feels compelled to appeal to something that *underlies* our language games: a mysterious property that stands behind—both in the sense of remaining invisibly in the background and in the sense of guaranteeing—our ordinary ways of speaking and acting. The metaphysical realist and the deflationist share a common picture in that it seems to both a queer thing that certain statements (for example, about the past) can be said to be true.

### The Error (and the Insight) in Verificationism

Part of what is right in the metaphysical realist’s response to the deflationist is the realization that that view does not (as advertised) successfully undercut Dummettian antirealism. On the contrary, deflationism about truth—as long as it involves (as it has since Paul Ramsey introduced the position in the 1920s) a verificationist account of understanding—adopts the most disastrous feature of the antirealist view, the very feature that brings about the loss of the world (and the past). It differs from antirealism in this regard only in that it attempts to disguise that feature

by means of a superficial terminological conservatism. The metaphysical realist is thus right to this extent: to undercut Dummett's antirealism requires challenging his account of understanding, not adopting it. But what makes the metaphysical realist's response *metaphysical* is its acceptance of the idea (which it shares with the Dummettian antirealist) that our ordinary realism—for example, about the past—presupposes a view of truth as a “substantive property.” The metaphysical realist, in wanting a property that he can ascribe to all and only true sentences, wants a property that corresponds to the assertoric force of a sentence. But this is a very funny property. To avoid identifying this property of “truth” with that of assertability, the metaphysical realist needs to argue that there is something we are saying when we say of a particular claim that it is true over and above what we are saying when we simply assert the claim. He wants truth to be something that *goes beyond* the content of the claim and to be that in virtue of which the claim is true. This forces the metaphysical realist to postulate that there is some single thing we are saying (over and above what we are claiming) whenever we make a truth claim, no matter what sort of statement we are discussing, no matter what the circumstances under which the statement is said to be true, and no matter what the pragmatic point of calling it true is said to be.<sup>12</sup>

The right alternative to thinking of truth as a “substantive property” à la the metaphysical realist is not to think of our statements as mere marks and noises that our community has taught us to associate with conditions for being conclusively verified (as in the account of Dummett's “global antirealist”), or to associate with “betting behavior” in a way that is “a function of observable circumstances” (as in Horwich's account). The right alternative is to recognize that empirical statements already make claims about the world—many different sorts of claims about the world—whether or not they contain the words ‘is true’. What is wrong in deflationism is that it cannot properly accommodate the truism that certain claims about the world are (not merely assertable or verifiable but) *true*. What is right in deflationism is that if I assert that “it is true that *p*,” then I assert the same thing as if I simply assert *p*. Our confidence, when we make statements about the past, that we are saying something whose rightness or wrongness depends on *how things were back then* (when we claim, for example, that “It is true that Lizzie Borden killed her

parents with an axe”) is not something that requires the metaphysical idea that there is a “substantive property” whose existence underwrites the very possibility of using the word ‘true’.

In order to see more clearly the difference between the commonsense realism I am defending and the kind of metaphysical realism we are right to recoil from, let us shift our attention for a moment from discourse about observable things, such as deer grazing on the meadow, to discourse about unobservables, for example, microbes. In the first lecture, I remarked that the use of instruments should be viewed as a way of extending our natural powers of observation. But the use of language is also a way of extending our natural powers of observation. If I could not understand talk about “things too small to see with the naked eye,” the microscope would be at best a toy (like the kaleidoscope); what I saw when I looked through the eyepiece would mean nothing to me. It would be a mistake, however, to conclude that the dependence goes both ways. The phrase ‘too small to see with the naked eye’ does not depend for its intelligibility on the invention of an instrument that allows us to see things smaller than the things that the naked eye can see (nor did we regard it as changing its sense when the microscope was invented). What is mistaken about verificationism is the claim that the meaning of an expression like ‘things too small to see with the naked eye’ depends on there being methods of verifying the existence of such things, and the related claim that the meaning of such an expression changes as these methods of verification change (for example, with the invention of the microscope). There is a philosophical danger, however, of rejecting what is right in verificationism in the course of rejecting what is wrong with it. What is right in verificationism is that a great deal of scientific talk does depend for its full intelligibility on the provision of the kind of thick explanatory detail that is impossible if one has no familiarity with the use of scientific instruments. For example, in Democritus’s writings, as we know of them, the notion of an “atom” was a metaphysical one, but one to which *we* can give a sense, even if Democritus himself could not.<sup>13</sup> Thus, scientific instruments and scientific ways of talking are both ways of extending our perceptual and conceptual powers, and those ways are highly interdependent; indeed, they can fuse into a single complex practice.

The ways in which language extends the mental abilities that we share with other animals are almost endless; our ability to construct sophisticated scientific theories is only one example. A very different sort of example is provided by the role of logical constants, for example, the words 'all' and 'no'. An animal or a child that has not yet learned to use these words may have expectations that we who have acquired them can and do describe with the aid of these words. For example, imagine that someone with modest skills at sleight-of-hand causes a handkerchief to "vanish" in front of a child's very eyes, and the child displays astonishment. We might say that the child believes (believed) that "handkerchiefs do not vanish into thin air just like that"—that is, that *no* handkerchiefs vanish into thin air just like that. Of course, that generalization does not have any consequences that the child can understand not possessed by the generalization: "*observed* handkerchiefs do not vanish into thin air just like that." Yet we would not dream of using the latter words to describe the child's attitude to the event. We would not know how to make sense of the suggestion that a child is only concerned to make a judgment about the behavior of *observed* handkerchiefs. This is the case *not* because we take the child to be concerned with making judgments about both observed *and* unobserved handkerchiefs; the distinction between the two generalizations is not one that belongs to the child's intellectual repertoire. It is a part of *our* repertoire (and which description we use may make a difference to *us* under certain circumstances: "Fine shades of behavior. Why are they *important*? They have important consequences"<sup>14</sup>). We describe even primitive preverbal attitudes as attitudes toward objects of which people may or may not be aware, and not just toward the part of the world that the child (or we) can "verify." Our sophisticated adult talk about certain features of the world (such as "those which are observable to us") rests upon—is parasitic upon—just such a primitive preverbal attitude toward the world.

A quite different aspect of the extension of our conceptual abilities brought about by the possession of words for generality is the possibility of formulating conjectures that transcend even "ideal verifiability," such as "There are no intelligent extraterrestrials." The fact that this conjecture may not be verifiable even "in principle" does not mean that it does

not correspond to a reality; but one can say what reality corresponds to it, if it is true, only by using the words themselves.<sup>15</sup> And this is not deflationism; on the contrary, deflationism, by identifying understanding with possession of verification abilities, makes it mysterious that we should find these words intelligible. Once again, the difficulty here lies in keeping what is right in verificationism (or in this case in deflationism) while throwing out what is wrong. . . .

### Wittgenstein on Truth

How, then, do we understand “recognition-transcendent” uses of the word ‘true’, as, for example, when we say that the sentence ‘Lizzie Borden killed her parents with an axe’ may well be true even though we may never be able to establish for certain that it is? Tarski (who was not a deflationist in my sense, because he never endorsed the verificationist account of understanding in any of its versions) expressed a genuine insight in pointing out (as Gottlob Frege had before him) that there is an intimate connection between understanding a sentence and understanding the claim that that sentence is true. If we accept it that understanding the sentence ‘Lizzie Borden killed her parents with an axe’ is not simply a matter of being able to recognize a verification in our own experience—accept it, that is, that we are able to conceive of how things that we cannot verify *were*—then it will not appear as “magical” or “mysterious” that we can understand the claim that that sentence is *true*. What makes it true, if it is, is simply that Lizzie Borden killed her parents with an axe.<sup>16</sup> The recognition transcendence of truth comes, in this case, to no more than the “recognition transcendence” of some killings. And did we ever think that all killers can be recognized as such? Or that the belief that there are certain determinate individuals who are or were killers and who cannot be detected as such by us is a belief in magical powers of the mind?

There is, however, something that Tarski ignores, and that is the fact that there are perfectly well-formed declarative sentences that are *neither* true nor false; indeed, in Tarski’s theory, it was supposed to be a theorem of logic (given what Tarski calls an “adequate definition” of the truth predicate<sup>17</sup>) that each sentence is either true or false (has a true negation).

But there are many reasons why a sentence may fail to have a truth value: for example, the vagueness of some of its terms ('The number of trees in Canada is even'), or the failure of the world to behave the way it should if the terms it employs are to work (for example, many sentences about the simultaneity of events were *discovered* to lack a truth value when relativity theory appeared on the scene; this is quite different from ordinary vagueness, of the kind that it requires only "linguistic intuition" to perceive). The use of 'true' and 'false' in "Such and such a sentence is *neither true nor false*" is inadmissible in Tarskian semantics. Those who regard 'true' as a mere "device for disquotation" (for example, asserting sentences without actually using them), also ignore or deny this clearly predicative use of 'true' and 'false'.

One thinker who did *not* ignore or deny this was Wittgenstein. In an important (but frequently misunderstood) section of *Philosophical Investigations*, he writes:

At bottom, giving "This is how things are" as the general form of propositions<sup>18</sup> is the same as giving the definition: a proposition is whatever can be true or false. For instead of "This is how things are" I could have said "This is true." (Or again "This is false.") But we have

'*p*' is true = *p*

'*p*' is false = not-*p*

And to say that a proposition is whatever can be true or false amounts to saying: we call something a proposition when in our language we apply the calculus of truth functions to it.

Now it looks as if the definition—a proposition is whatever can be true or false—determined what a proposition was, by saying: what fits the concept 'true', or whatever the concept 'true' fits, is a proposition. So it is as if we had a concept of true and false which we could use to determine what is and what is not a proposition. What engages with the concept of truth (as with a cogwheel) is a proposition.

But this is a bad picture. It is as if one were to say "The king in chess is the piece that one can check." But this can mean no more than that in our game of chess we only check the king. Just as the proposition that only a proposition can be true or false can say no more than that we only predicate "true" and "false" of what we call a proposition. And what a proposition is is in one sense determined by the rules of sentence formation (in English, for example), and in another sense by the use of the sign in the language-game. And the use of the words "true" and "false" may be among the constituent parts of the game; and if so it belongs to our concept 'proposition' but does not 'fit' it. As we might also say, check belongs to our concept of the king in chess (as so to speak a constituent part of it). To say

that check did not fit our concept of the pawns, would mean that a game in which pawns were checked, in which, say, the players who lost their pawns lost, would be uninteresting or stupid or too complicated or something of the kind (§136).

Kripke, who quotes only “But we have ‘ $p$ ’ is true =  $p$ ,” sees §136 as a clear expression of deflationism. But, for the following reasons, I do not believe this can be what Wittgenstein intended.

1. We know that Wittgenstein does not oppose the idea that empirical propositions “correspond to realities”; indeed, he elsewhere discusses the sense of this correspondence, and distinguishes it from the very different sense in which mathematical propositions correspond to reality;<sup>19</sup> rather, the thrust of the whole passage is clearly directed against the metaphysical realist’s understanding of such platitudinous thoughts as the thought that “This chair is blue” can correspond to the fact that a particular chair is blue. The essential point Wittgenstein makes in §136 is that we do not recognize that something is a proposition by seeing that it “fits” the concept “truth,” where truth is conceived of as a free-standing property.<sup>20</sup> But it would be exactly as much of a mistake to think that we can explain what truth is by saying that for any *proposition*  $p$ ,  $p$  is true =  $p$ , as it is to think that we can explain what a proposition is by saying that a proposition is what is true or false. In both cases, we are simply making grammatical observations; we must not confuse what are virtually tautologies for metaphysical discoveries. The notion of truth and the notion of a proposition mesh together like a pair of gears in a machine; neither is a foundation on which the other rests. Our understanding of what truth comes to, in any particular case (and it can come to very different things), is given by our understanding of the proposition, and that is dependent on our mastery of “the language game,” by which Wittgenstein means here “the whole, consisting of language and the actions into which it is woven.”<sup>21</sup> There is a certain “holism” here; knowing what truth is in a particular case depends on knowing the *use* of signs in the language game just as knowing what checking is depends on knowing the use of the various pieces in chess.

2. When we *ourselves* are willing to apply truth functions to a sentence—note how Wittgenstein emphasizes *in our language*—we regard the sentence as true or false, as a genuine *Satz*.

3. A grammatical string of sounds or marks which is *neither* true nor false is simply not a sentence (*Satz*) in Wittgenstein’s sense.<sup>22</sup> This is what Wittgenstein means by speaking of “*the definition*—a proposition is whatever can be true or false” [my emphasis]. There is no suggestion in this that adding the words ‘is true’ is a “logical device” that we can apply to “declarative sentences” *ad libitum*.<sup>23</sup>

The possibility that I see in Wittgenstein's writings, of doing full justice to the principle that to call a proposition true is equivalent to asserting the proposition (doing full justice to what I called "Tarski's insight") without committing the errors of the deflationists, is a condition of preserving our common-sense realism while appreciating the enormous *difference* between that common-sense realism and the elaborate metaphysical fantasy that is traditional realism—the fantasy of imagining that the form of all knowledge claims is fixed once and for all in advance. That fantasy goes with the equally fantastic idea that there must be just one *way* in which a knowledge claim can be responsible to reality—by "corresponding" to it, where "correspondence" is thought of as a mysterious relation that somehow underwrites the very possibility of there being knowledge claims. Indeed, a rejection of the idea that we can speak once and for all of "all propositions" as if these constituted a determinate and surveyable totality, and of one single "truth predicate," whose meaning is fixed once and for all, is also one that the later Wittgenstein shared with Tarski.<sup>24</sup>

Instead of looking for a free-standing property of "truth" in the hope that when we find what the property is we shall know what the *nature* of propositions is and what the *nature* of their correspondence to reality is, Wittgenstein wants us to *look* at ethical language (and not the kind of ethical language that only occurs in philosophy<sup>25</sup>), to look at religious language,<sup>26</sup> to look at mathematical language, which is itself, he says, a "motley,"<sup>27</sup> to look at imprecise language that manages to be perfectly "clear" in context ("Stand roughly here"<sup>28</sup>), to look at talk that is sometimes nonsensical and to look at the very same sentences when they function perfectly well (talk of "what is going on in so-and-so's head" is an example of this<sup>29</sup>), to look and *see* the differences in the way these sorts of discourse function, all the very different ways in which they relate to reality.

If Wittgenstein was right, how should his reflections affect our view of the concept of truth? On the one hand, to regard an assertion or a belief or a thought as true or false *is* to regard it as being right or wrong; on the other hand, just what sort of rightness or wrongness is in question varies enormously with the *sort* of discourse. 'Statement', 'true', 'refers', indeed 'belief', 'assertion', 'thought', 'language'—all the terms that we use when



we think about logic (or “grammar”) in the wide sense in which Wittgenstein understands that notion—have a plurality of uses, and new uses are constantly added as new forms of discourse come into existence. On the other hand, that does not mean that any practices at all of employing “marks and noises” can be recognized by us as adding up to a form of discourse—for not every way of producing marks and noises is “one in which there is the face of meaning at all.”<sup>30</sup> Part of what I have been trying to show in these lectures is that what we recognize as the face of meaning is, in a number of fundamentally important cases, also the face of our natural cognitive relations to the world—the face of perceiving, of imagining, of expecting, of remembering, and so on—even though it is also the case that as language extends those natural cognitive relations to the world, it also transforms them. Our journey has brought us back to the familiar: truth is sometimes recognition-transcendent because what goes on in the world is sometimes beyond our power to recognize, even when it is not beyond our power to conceive.

#### Notes

This chapter has been excerpted, with omissions, from Putnam’s 1994 Dewey Lectures, Lecture 3, “The Face of Cognition,” *Journal of Philosophy* 91, no. 9.

1. *The Logical Basis of Metaphysics* (Cambridge: Harvard, 1991).
2. “The Concept of Truth in Formalized Languages,” reprinted in his *Logic, Semantics, Metamathematics* (New York: Oxford, 1956).
3. If the sentence *S* is not in English, then we must write the translation of the sentence *S* into English in the blank.
4. Jan Wolenski, a scholar who has spent a many years studying the history of Polish logic and philosophy, has informed me that at the time Tarski wrote “The Concept of Truth in Formalized Languages” he held that nothing much could be said about what understanding a sentence consists in. The idea that Tarski agreed with logical-positivist accounts of language is just wrong, according to Wolenski. In “The Concept of Truth” itself, Tarski employs the notion of “ascribing concrete, and, for us, intelligible meanings to the signs” quite uncritically, (See *Logic, Semantics, Metamathematics*, pp. 166–7; see also Wolenski, “Tarski as a Philosopher,” *Poznan Studies in the Philosophy of the Sciences and the Humanities*, xxviii (1993): 318–38.)
5. Paul Horwich, a well-known deflationist, sums up the position thus in a recent review: “[I]t is a mistake to think that truth is a substantive property with some unified underlying nature awaiting philosophical articulation. Rather, our truth predicate is merely a logical device enabling simple formulations of certain kinds

of generalizations ... and the concept of truth is entirely captured by stipulating the equivalence schema, 'The proposition that  $p$  is true if and only if  $p$ '—where  $p$  can be replaced by any declarative sentence"—"In the Truth Domain" (a review of Crispin Wright, *Truth and Objectivity*), *Times Literary Supplement* (July 16, 1993), p. 28.

6. See, for example, Horwich, "Wittgenstein and Kripke on the Nature of Meaning," *Mind and Language*, v, 2 (Summer 1990): 105–21. Horwich writes, "The communal disposition to use a word in a particular way should not be regarded as simply the disposition to treat certain sentences as definitely and permanently acceptable and others not. In addition, there are dispositions to sanction various levels of confidence (*cashd out as 'betting behavior'*) in the truth of certain sentences—where the appropriate degrees of belief are *a function of observable circumstances*" (p. 112, emphasis added). Horwich has published a book-length defense of deflationism titled *Truth* (Cambridge: Blackwell, 1990). Note that when Horwich says that this theory is not "committed to verificationism" (p. 114), all he means is that he is prepared to say that a sentence can be said to be true or false even if its verification conditions do not determine that it is "determinately" true or false—indeed, this follows from the decision to retain the principle of bivalence as a logical truth; Horwich's account of *what understanding consists in* is precisely Carnap's, down to the identification of confidence with "betting behavior."

7. According to Horwich ("In the Truth Domain"), one of the purposes of what he calls "the truth predicate" is to enable us to make the generalization "All propositions of the form ' $p$  or not  $p$ ' are true," where  $p$  is "any declarative sentence." In particular, even if  $p$  is a sentence whose truth value we might consider indeterminate—for example, "A broken chair is still a chair"—logic forces us to say that the sentence *is* true or false, on Horwich's account; cf. *Truth*, pp. 80–8.

8. If I distinguish here between Dummett himself and the position he calls "global antirealism," it is because Dummett himself frequently expresses some dissatisfaction with the counterintuitiveness of global antirealism, and some uncertainty as to its correctness. But, I would argue, it is because he structures the debate in the way I describe that he sees no satisfactory alternative to global antirealism.

9. Cf. Carnap. "Truth and Confirmation," in *Readings in Philosophical Analysis*, H. Feigl and W. Sellars, eds. (New York: Appleton-Century-Crofts, 1949), for as clear and forceful a statement of what has come to be called "deflationism" as has ever been given.

10. As pointed out in footnote 5. Horwich believes that there are "substantive" things to be said about degrees of warranted assertability—for example, that they are determined, at least loosely, by "communal standards," and that they establish legitimate "degrees of confidence" which are in turn to be interpreted as "betting behavior."

11. *Philosophical Investigations* §195.

12. I fell into this error myself in my previous published criticisms of deflationism ("On Truth" and "Does the Disquotational Theory of Truth Solve All Philosophical Problems?" both reprinted in *Words and Life*).

13. In this connection, see Cora Diamond's discussion of the "sense" of riddles, in "Riddles and Anselm's Riddle," *Proceedings of the Aristotelian Society*, supp. vol. II (1977), reprinted in her *The Realistic Spirit* (Cambridge: MIT, 1991).

14. Wittgenstein, *Philosophical Investigations*, p. 204.

15. I borrow this use of 'correspond to a reality' from Wittgenstein's *Lectures on the Philosophy of Mathematics*, Diamond, ed. (Chicago: University Press, 1989), lecture 25.

16. Note that from the fact that a "that clause" is a *nominalization*, it does not follow that we have to postulate an *object* that it names. Davidson (who is following Tarski here) is right in maintaining that the connection between the "fact," if it is a fact, that Lizzie Borden committed the famous murder (or whatever the example in question may be) and the truth of the sentence we are using as an example can be stated as a simple biconditional: "Lizzie Borden did commit the famous murder" is true if and only if Lizzie Borden did commit the famous murder; and that biconditional does not contain a "that-clause." Even sentences with (apparently) ineliminable that-clauses, for example, "John believes that Lizzie Borden did commit the famous murder," do not have to be interpreted as asserting a *relation between a belief and a proposition* (contrary to the view of Fodor). The tendency to postulate entities whenever one finds quantifiers used is the legacy of Quine's "criterion of ontological commitment"; my reasons for rejecting the whole idea of such a criterion are briefly stated in Lecture I, footnote 12 [of "Sense, Nonsense, and the Senses: An Inquiry into the Powers of the Human Mind," *Journal of philosophy* 91, no: 9.—Ed.]

17. In "A Comparison of Something with Something Else," in *Words and Life*, I argue that Tarski's so-called "truth definitions" are at best *extensionally correct*, they do not yield correct characterizations of truth under counterfactual circumstances, and they certainly do not tell us what 'true' *means*.

18. "This is how things are" was given as the general form of propositions in Wittgenstein's *Tractatus Logico-Philosophicus*.

19. See, for example, Wittgenstein's distinction of two very different notions of "corresponding to reality" in *Lectures on the Philosophy of Mathematics*, lectures 25 and 26. Among other things. Wittgenstein says that "This chair is blue" (imagine he had a blue chair in front of him) corresponds to a reality, but he can only say to what reality by using the sentence itself. He also says that while the sentences of arithmetic do not correspond to a reality in *that* sense, the *practice* of arithmetic does, in a different sense, correspond to "a diffuse empirical reality."

20. Notice that a little later, in §138, Wittgenstein *rejects* the idea that "the meaning of a word I understand fit[s] the sense of a sentence I understand," saying "Of course, if the meaning is the *use* we make of the word, it makes no sense to speak of such 'fitting'." A similar contrast between thinking of meaning as use

and thinking of the possibilities of use as fixed by the ways in which meanings “fit” or fail to “fit” one another in already drawn in Wittgenstein’s 1932–35 Lectures. Wittgenstein’s rejects the idea that we can explain what a proposition is by appeal to the notion of fitting the sense of “true” *because he rejects the idea of “fitting” involved*, not because he is offering a philosophical thesis about the meaning of “true.”

21. Cf. *Philosophical Investigations* §7.

22. As I pointed out earlier, Wittgenstein thinks of a “sentence” (*Satz*, translated as “proposition” by Anscombe) *neither* as a sentence in the sense in which logicians speaks of “sentences,” that is a mere string of marks or noises, *nor* as a “proposition” in the sense in which some philosophers do, that is, as a “sense” (in abstraction from the sign-design that carries that sense). Wittgenstein rejects that kind of “sentence/proposition” distinction. Deflationists read the formula “*p* is true = *p*” as meaning that to produce the mark or noise *p* *is true* is equivalent to producing the mark or noise *p*, but Wittgenstein is not talking about writing *marks* or producing *noises*.

23. See Horwich’s formulation of deflationism quoted in footnote 5.

24. Although Tarski never pretended to be a philosopher of language, his profound logical investigation into the liar paradox and the other so-called “semantical paradoxes” (investigations which built on the techniques Gödel used to prove the celebrated incompleteness theorem) convinced him that, on pain of paradox, we may only regard “true” as a well-defined concept when that predicated is restricted to a single “language,” a single determinate totality of propositions, and that the judgment that a member of the totality is itself true or false may not belong to the totality on pain of contradiction. An immediate corollary of this Tarskian view is that the totality of possible propositions is inherently unsurveyable. For details, see the discussion of the liar paradox, and the version known as the “strong liar,” in the title essay of *Realism with a Human Face*. Today, not all logicians agree with Tarski that a consistent language may never contain its own truth predicate; but the “non-Tarskian” ways of avoiding the Liar paradox that have been proposed by Kripke and others still have the property that the semantics of a consistent language cannot be completely given *in* the language itself. As Kripke has put it, “The ghost of the Tarskian hierarchy is still with us.” Of course, Wittgenstein’s reasons for regarding language as an ever growing body of ways of speaking and thinking, with an unpredictable variety of ways of “corresponding to reality,” do not have to do with the problem of the formal antinomies that concerned Tarski.

25. Rush Rhees, who understood Wittgenstein’s philosophy as well as anyone, once wrote “If anyone does ask “What are moral statements like?” I should think one ought to begin by giving examples of them. But often writers on ethics do not do this. You mention “Honesty is good.” I cannot remember ever hearing anyone say this, unless it be in a philosophy discussion. And I cannot imagine just the circumstances under which anyone *would* say it”—*Without Answers* (New York: Routledge, 1969), p. 103.

26. For a discussion of Wittgenstein on religious language, see *Renewing Philosophy*. chs. 7, 8.

27. *Remarks on the Foundations of Mathematics* §46 (“A motley of techniques of proof”) and §48 (“I want to give an account of the motley of mathematics”).

28. Note that the reason this is clear in context is certainly not that the context makes it perfectly precise! It is that exactness has no place here—and, as Wittgenstein says (*Philosophical Investigations* §69) “you still owe me a definition of exactness.”

29. *Philosophical Investigations* §422 to §427.

30. Diamond, *The Realistic Spirit*, p. 261

---

## A Functionalist Theory of Truth

Michael P. Lynch

To copy a reality is, indeed, one very important way of agreeing with it, but it is far from being essential.

William James

The question is not whether [“true” and “false”] are in practice applied to ethical statements, but whether, if they are so applied, the point of doing so would be the same as the point of applying them to statements of other kinds, and if not, in what ways it would be different.

Michael Dummett

The history of attempts to identify the property that all and only true propositions have in common has not been a happy one. Traditional theories that aim to provide us with the essence of truth—correspondence, coherence, pragmatist, and so on—each face well-known difficulties that prevent their advocates from striking their tents and declaring victory. Indeed, it is just these problems that deflationists point to when arguing that it is best to simply abandon robust theories and admit that truth has no underlying nature.<sup>1</sup>

This lack of success in explaining the nature of truth is doubtless due to a number of causes, not the least of which is the difficulty of the subject. But when I look back over the various attempts—many of which are represented in this volume—I see a particular pattern of failure. Baldly put, that pattern goes something like this. A theory of truth is proposed and argued for by appeal to propositions of a certain domain, the truth of which the theory seems to explain quite nicely. The theory is then extended to cover propositions of every domain. But this extension runs up against counterexamples; that is, the theory does not explain how propositions from certain domains can be true.

Consider how this problem of scope arises for a popular form of the correspondence theory. That theory understands the correspondence relation in terms of the referential/causal properties of a proposition's constituents.<sup>2</sup> The proposition that the book is on the table is true in virtue of its components being causally/referentially related to certain mind-independent objects—a particular book and a particular table. But as plausible as this account may be when applied to propositions about middle-sized dry goods, it is much less plausible when applied to propositions about numbers, such as the proposition that the number six is even. Whatever else numbers might be, they presumably are not physical objects. No number is ever in causal contact with our thought. Thus our thought that the number six is even can't be true in virtue of a causal relationship between its components and the number six. Or consider propositions like those we hear on the nightly news about the status of the economy or the constitutionality of particular laws. Some of these propositions are true. But economies and laws are dubious candidates for physical, causally efficacious objects. Further, far from being mind-independent, economies and laws are paradigmatically conventional. Thus it is a puzzle how propositions about them can be understood in terms of correspondence with mind-independent objects.

I bring up the causal theory of correspondence only as an example. In my view, similar problems of scope await other traditional theories of truth as well.<sup>3</sup> Epistemic theories are notoriously unable to explain the truth of propositions about humanly inaccessible parts of the universe or about the past. A theory that defines truth in terms of what would be justifiedly believed cannot explain how the proposition that *the number of the stars in the universe right now is odd* can even have a truth value. Even the deflationary view of truth faces this problem of scope. This may be a bit surprising because one might think that deflationism would not have a problem explaining the metaphysical nature of truth. After all, the core of the position is that truth has no nature. Yet this itself is a metaphysical position (see Devitt, chap. 25), and if deflationism is to succeed as its advocates wish, it must be equally plausible in every domain. Yet as several essays in this volume reveal, deflationary views face their own problems explaining our commitment to certain classes of propositions—blind ascriptions of truth, generalizations involving the concept of

truth, and indeterminately true propositions, to name just three examples (see Gupta, chap. 23, and Field 1997).<sup>4</sup>

Proponents of both robust and deflationary theories will protest that there are responses available to these problems. And, of course, there are. I shall not argue here that each set of counterexamples is insurmountable. Rather, I note that two facts stand out when looking at all these problems together. First, each traditional theory of truth is more plausible in some domains than in others. Second, all of the theories mentioned are assuming that the question “What is truth?” has a single answer. In other words, most of the players in the contemporary debate over truth share an unnoticed allegiance to a certain type of monism: truth has but one underlying nature—if any nature at all.

Once these facts are noticed, it is hard not to see them as related. Therefore, it seems relevant to reconsider our commitment to this alethic monism. At the very least, it seems important to think through possible alternatives. A further reason for pursuing this course is that the monism that comes so naturally to philosophers runs contrary to the way most other people think about truth. Most folks think that different sorts of propositions can be true without being true in the same way. Their intuitive thought is that moral or legal propositions can be true all right; it is just that their truth is of a different kind than that of propositions like the cat is on the mat. Such considerations are not decisive, but they suggest that alethic monism is an artifact of philosophical theory, not a result of ordinary practice.

Until recently, discussions that questioned alethic monism were absent from the contemporary literature. But the subject has been broached by several philosophers in the last few years, most notably by Crispin Wright: “The proposal is simply that any predicate that exhibits certain very general features qualifies, just on that account, as a truth predicate. That is quite consistent . . . with acknowledging that there is a prospect of pluralism—that the more there is to say may well vary from discourse to discourse” (Wright, 1992, p. 38). The general features Wright is talking about here are what he calls “platitudes,” or a set of a priori principles, about the concept of truth. These include the principles that the proposition that *p* is true if and only if *p*, that truth is distinct from justification, that truth is timeless, and so on. In his original statement of the view,



Wright argued that any predicate that satisfies a minimal set of these platitudes qualifies as a truth predicate. But in some discourses, the truth predicate may have more robust content, depending on the types of additional platitudes about truth that the discourse brings along in its wake. In short, Wright's view allows for the possibility that there may be different kinds of truth.

In recent essays, both Hilary Putnam and Terry Horgan have also entertained this idea (e.g., Putnam, chap. 30; Horgan, chap. 4).<sup>5</sup> Most philosophers, however, continue to look askance at the suggestion that truth itself might come in different kinds. Different kinds of truths (propositions), yes; different kinds of truth, no. The prevailing thought seems to be that even if the nonphilosopher might sometimes *talk* as if she thought moral truth was of a different kind than physical truth, this is not a claim we should take seriously; such talk is just that: talk.

One reason for this skepticism is that a plurality of kinds of truth seems to imply a plurality of truth concepts. And a plurality of truth concepts entails that the word "true" is ambiguous.

The worry is a real one. The ambiguity of truth would have several ill consequences. First, it would undermine one of the most useful and important functions of the truth concept. As is often remarked, our concept of truth allows us to generalize "blindly" over propositions of all sorts, as in "Everything Socrates said was true." If the word "true" were ambiguous in the way that "bank" or "step" happen to be—if its meaning completely changes from context to context—then attempts to ascribe truth to everything Socrates ever said, no matter what the context, would be impossible. Yet clearly we can generalize over propositions in this way, and hence truth must not be ambiguous.

Second, the equivocality of truth would make validity and logical inference mysterious. Instances of valid argument forms whose premises are mixed, i.e., whose premises stem from different discourses, would end up equivocating (see Sainsbury 1996). Consider a simple argument:

If violence causes pain, then it is wrong.  
Violence does cause pain.  
Therefore, violence is wrong.

This is obviously valid; its premises are truth-preserving. Yet its second premise is a descriptive statement about cause and effect relations in the world, while its conclusion is a moral evaluation. Were the ambiguity thesis to hold with regard truth, these statements could be true in completely different senses; they could literally have different properties, each expressible by the predicate “true.” The argument might no longer “preserve” a single property at all.

In my view, these considerations suggest that our semantic or conceptual account of truth must be uniform across context.<sup>6</sup> But it does not imply, I shall argue, that our account of the deep nature of truth must be similarly uniform. We can be monists about the concept of truth while being pluralists about its underlying nature. The key is to see the concept of truth as the concept of a multiply realizable property.<sup>7</sup>

## 1 Truth as a Functional Concept: A First Pass

In the philosophy of mind, it is common to the point of dogma to say that mental states are functional states. A particular mental state such as pain, or the belief that it will rain, is the state that it is in virtue of the functional role it has within our overall cognitive system. But what realizes or plays that role in any particular organism may vary: mental states are therefore said to be “multiply realizable.”

The terms “function” and “functional role” mean different things to different people.<sup>8</sup> In the general and intuitive sense I am concerned with, a function is a type of job. To occupy a functional role is therefore to satisfy a certain job description. Consequently, a *functional concept* is the concept of a property, state, or object that occupies or plays such a role. So for example, being a head of state, being a carburetor, and being a heart are all functional concepts in my sense. A heart is anything that has the function of pumping the blood throughout the body of an organism, and various structures, organic and inorganic, might play or “realize” this role within some particular cardiovascular system. Indeed, the underlying nature of what plays the heart role in a given system can be quite different in different organisms. But it is the underlying nature of whatever plays that role that explains the performance of that function.

As an intuition pump, consider one of our most basic platitudes about truth: a proposition is true when the world is as that proposition says that it is. This platitude (and its more formal brethren, the disquotational and equivalence schemata) captures something about the very essence of truth. Yet paradoxically, it tells us nothing specific about that essence. Robust theorists take platitudes of this sort as the starting point in any theory of truth, a minimum requirement that every theory must meet, while deflationists take it to be the end of any such theory. Both attitudes overlook a subtly different way of understanding the main lesson of the platitude. Rather than attempting to read it as the key to truth's deep metaphysical nature (or lack thereof), we can understand our platitude as specifying truth's "job" in our conceptual scheme—its *functional role*. We can understand it as telling us that our concept of truth is the concept of *whatever* property a proposition has when the world is as that proposition says that it is. Roughly speaking, "saying it like it is" is part of the functional role of true propositions, and propositions that do so (and also fulfill various other conditions, as we shall see) have the property of truth.

Consider another analogy.<sup>9</sup> The position of head of state is found in almost every constituted government. It is held by presidents, prime ministers, kings, queens and even religious figures, all of whom are head of state in virtue of performing a certain job, namely, being the chief executive officer for the government. This job will vary from country to country in some respects (although it will remain constant in others), and individual heads of state may have properties that others do not have (such as being elected, which is true of presidents and popes but not of most kings). And yet when we say that both Fidel Castro and Bill Clinton are heads of state, we are not equivocating. They simply perform or realize the functional role of being a head of state differently. In the same sense, for the functionalist about truth, truth talk is not equivocal, because the *concept* of truth is everywhere understood as naming a particular functional role. Nonetheless, what realizes that role may vary from context to context.

Much more needs to be said, of course, to make sense of this idea. But we can already see that a functionalist theory of truth will leave room for the possibility that in the case of propositions about middle-sized dry

goods, the claim that *a* is *F* may be true in virtue of the referential properties of that claim. Causally understood correspondence relations might realize the functional role of truth in such cases. Yet the functionalist account could also allow a different property—coherence with other propositions, for instance—to occupy the role of saying how things are in other contexts. Functionalism about truth is therefore consistent with alethic pluralism at the deep metaphysical level: the underlying nature of what realizes truth in any particular mode of thought is still an open question.

If the concept of truth can be seen as naming a functional role, we should be able to specify that role in an informative way. Doing so is the goal of the next section.

## 2 The Alethic Network and the Role of Truth

According to what is often called “commonsense” functionalism, mental concepts such as belief and desire come together in a package deal. Like the concepts of soldier/army and teacher/student, they apply together or not at all (Armstrong 1999, 84). These concepts are not individuated one by one but by their place in the network of interrelated psychological generalizations that make up our commonsense psychology, or “the psychological platitudes which are common knowledge among us—every one knows them, everyone knows that everyone else knows them, and so on” (Lewis 1972, 208). These and an indefinite number of other such platitudes jointly carve out the causal roles of our mental states. If we take the mental state of pain, for example, the relevant platitudes will include such chestnuts as “People who are in great pain are usually not happy about it,” “One is typically aware of one’s pain,” and “A threat of pain typically causes fear.” String such platitudes together and we have a job description for pain, namely, that pain is a state that is typically causally related to certain inputs, outputs, and other mental states. The *concept* of pain will therefore apply to any property or state that fits this job description, that realizes this causal role. As Lewis says, a mental concept is “the concept of a member of a system of states that together more or less realize the pattern of causal generalizations set forth in commonsense psychology” (1980, 112). So the concept of pain, on this

account, is the concept of the state that more or less realizes the role of pain in our psychology, and whatever state does realize that role is pain.

A similar model helps us understand our concept of truth. Like our psychological concepts, what we might call our alethic concepts, or concepts like truth, fact, proposition, and reference, package-deal concepts. It has long been recognized that concepts like proposition and fact are definable only in terms of each other: a proposition is whatever is true or false, a fact is what makes propositions true, true propositions fit the facts, and so on. The fact that these sorts of concepts form such a tight-knit family is part of what needs to be explained by any theory of truth. Adopting a functionalist perspective on these concepts does just that. Just as our psychological concepts are definable in terms of their roles in a network of interrelated psychological platitudes, so the common-sense set of principles and platitudes that together constitute our having a robust sense of the true and the real form the *alethic network*. Some of the most central of these principles are the following:

- The proposition that *p* is true if and only if *p*.
- The proposition that *p* is false if and only if it is not the case that *p*.
- Propositions are what are true and false.
- Every proposition has a negation.
- A proposition can be justified but not true, and true but not justified.
- True propositions represent, or correspond to, the facts, and false ones do not.
- Facts are what make propositions true.

Crispin Wright (chap. 32) suggests that platitudes of this sort can be taken to compose what he calls an “analytic theory” of truth. On his account, these principles jointly determine the meaning of the word “true” without providing necessary and sufficient conditions for the application of the term.

Like Wright, the functionalist sees such principles as constituting our grasp of the concept of truth. But there are also some interesting differences between Wright’s view and an explicitly functionalist theory. First, it is important to see that on the functionalist account, these and other alethic principles aren’t simply a list. They form a structure. In the philosophy of mind, the relevant principles can be divided into those that

primarily concern the inputs and outputs of the system (typically, sensations and behavior, respectively), and those that concern relations between the mental states themselves. Not all of the principles involved need fall neatly into one category or another, although most will, and not all mental states need be functionally defined, although most will. Similarly, some alethic platitudes will primarily concern the relationship between truth and other closely connected core notions (fact, proposition), while some relate these concepts to those that are intuitively farther outside these core concepts. As an example of the latter, consider the following:

- To claim that *p* is true implies that one believes that *p*.
- One knows that *p* only if it is true that *p*.
- Honest people typically speak the truth.
- Deliberately asserting what you know to be false is a lie.

Relative to core platitudes concerning truth and fact, principles like the above serve to relate the core concepts to others in our overall cognitive system. Again, it need not be entirely clear how “central” a platitude is, or whether it primarily concerns relating alethic concepts to nonalethic concepts. Here, like everywhere else, types of concepts shade off into one another. And while central alethic concepts will be functionally understood, it will remain an open question whether concepts further removed from the center (concepts like knowledge, for example) will also be functionally defined.

In the case of human psychology, most of the platitudes will be causal in nature (e.g., “Pain causes worry”). But not all will be. Others, like “A toothache is a type of pain,” will be quasi-logical. With regard to alethic terms, this order is reversed: one would suspect that most of the alethic principles would be quasi-logical, although there is nothing to rule out the possibility that some may also be causal.<sup>10</sup>

If we grant the existence of the alethic network as a whole, we can characterize each of the alethic concepts in terms of the role it plays within the network. This is a second way in which the functional account of truth differs from Wright’s analytic theory. Since truth, fact, proposition, etc., are explicitly package-deal concepts according to functionalism, the same platitudes that demarcate our concept of truth will also demarcate our other alethic concepts. In short, we can take a core alethic

concept as the member of a system of properties that together realize the interlinked structure of platitudes partially set forth above. Thus truth, for example, can be seen as whatever property plays the role demarcated by the associated principles and platitudes—the truth role, in other words. Other central members of that system will similarly be functionally understood. Propositions are whatever is true or false, whatever is asserted or denied, that which can take the place of the unquoted variable in “*p* says that *p*,” and so on. Similarly, we can understand a *fact* as whatever makes a proposition true, as conforming to the schema “It is a fact that *p* if and only if *p*” etc.

The functionalist can spell this out in greater detail, and with much more rigor, by employing F. P. Ramsey’s and David Lewis’s method for defining theoretical terms (see Lewis 1970). This method treats the term (or concept) being defined *as if* it were a new term being introduced into an ongoing scientific theory. To employ the method here, we would first form a conjunction of all, or some weighted subset, of the commonsense or “folk” platitudes about truth and related concepts in our network, *A*. This conjunction will include both *T* terms and *O* terms. *T* terms are the names for properties with unspecified natures that the theory is being used to introduce or define, in this case, terms like “true” or “fact.” *O* terms are the terms in the “old,” introducing vocabulary, such as “cause,” “object,” “person,” “snow,” and so on. To make things easier, we’ll stipulate that each *T* predicate explicitly represents a property (so that, e.g., “*x* is true” becomes “*x* has the property of being true”). We can then imagine *A* being written down in one long sentence roughly like this:

$$A(T_1, \dots, T_n, O_1, \dots, O_n)$$

The next step is to replace the *T* terms (but not the *O* terms) with variables, and then to prefix an existential quantifier for each variable. The result is the modified “Ramsey sentence” of *A*. If “*t*<sub>1</sub>” is a variable standing in for “true,” then the modified Ramsey sentence of *A* could be used to define “true” roughly as follows:

$$(FT) \ x \text{ is true} \Leftrightarrow_{df} \exists t_1, \dots, \exists t_n [A(t_1, \dots, t_n, O_1, \dots, O_n) \ \& \ x \text{ has } t_1]$$

This says that *x* (a proposition, say) is true just when there are certain alethic properties *t*<sub>1</sub>, ..., *t*<sub>*n*</sub> that are related among themselves as well to

nonalethic properties as specified in **A** and “*x* has *t*<sub>1</sub>.” By functionally defining the word “true,” (FT) gives us a way of understanding the concept of truth according to which a proposition is true just in case it has a property that plays the truth role marked out in **A**. Further, (FT) reveals that the identity of every alethic concept depends on its relation to every other concept; in a sense, we define each concept not one by one but all at once, or *en masse*. That is, we define every alethic concept in terms of having a property that uniquely bears certain relations to the other properties expressed by our alethic concepts and to the referents of the *O* terms.

So, to be true is to play the truth role. But functionalism allows that this role might be realized or occupied by different properties. We can say that a property *realizes the truth role for a discourse* just when it is the unique realizer (or near perfect realizer; see below) of that role for the propositions that compose the discourse. Thus propositions from two different discourses may have distinct properties that realize the truth role. Naturally enough, how one understands the nature of a particular discourse will determine which property, if any, one takes as uniquely playing the truth role for *that* discourse. For example, if one understands moral propositions to be about the subjective feelings of the speaker, then this metaphysical position will rule out moral propositions’ being true in virtue of their relation to mind-independent facts. As we will shortly see with the case of legal discourse, one’s overall metaphysical views provide a collection of a priori constraints or conditions on the realization of truth for that discourse. But functionalism itself remains neutral on just what these constraints will be.

We should not expect there to be a sharp and clear line between discourses or forms of thought. In ordinary life, we know the difference between talking about physics and talking about ethics. But sometimes we may say things that don’t clearly fall into either category, as we might if we were involved in a discussion about the ethical consequences of certain physical experiments for instance. It follows that we should expect some vagueness as to what discourse a particular proposition belongs. But I don’t need to know what discourse I am engaging in to know what I mean when I say that something is true. According to the functionalist view, it is a fact about the *concept* of truth that no matter what discourse



I may happen to be engaged in, what I say will be true just when it has a property that plays the truth role for that discourse. If what I say, the proposition I express, is not clearly a member of one discourse or another, how its truth is realized will also be unclear. But that is a question for deep metaphysics—not something I need to know in order to understand what I mean by the word “true.”

A functional characterization of the alethic concepts fits well with our holist intuitions about concepts like truth and fact. Because of their obviously interconnected nature, we learn such concepts not one by one but more or less all at once, as the functionalist theory implies. This is really no more mysterious than learning a complex skill like riding a bicycle (Heil 1998, 102). Acquiring this skill requires the coordination of a whole set of tinier skills—balancing, pedaling, steering, etc.—none of which are learned in isolation. You learn to master one as you come to master another, and vice versa. The alethic concepts are similar in this respect. Here as elsewhere, “light dawns gradually over the whole” (Wittgenstein 1969, sec. 141).

So far we’ve seen how (FT) explicates the concept of truth. But we also want to know what truth *is*, or about the *property* of truth. Those familiar with the debates over psychological functionalism know that we face a choice at this point. Functionalist theories in the philosophy of mind differ over whether to take mental properties as identical to the role properties or the realizer properties. The first alternative sees any given mental property as a “higher-order” property, or the property of having a property that plays a certain causal role; the second identifies properties with the “first-order” properties, or properties that realize that causal role in a system. In the same vein, we must choose between saying that the property of truth is the higher-order property of having a property that plays the truth role and saying that it is identical to the lower-order property that realizes that role in a particular discourse.

There are two reasons to prefer the former alternative when it comes to truth. First, to identify truth with its realizer property in some context (coherence, correspondence or whatever it might turn out to be) would immediately raise the ugly problem of ambiguity. Should it turn out that the truth role is realized by different properties in different contexts, we would no longer be able to talk of truth simpliciter but only of moral

truth, mathematical truth, physical truth, and so on. Second, truth is our chief cognitive goal. We want our beliefs to be coherent, or to correspond to fact, *because we want them to be true*, not the other way around. But this means that in cognition we are aiming at truth itself, not the properties that exemplify or realize it. Therefore, we should take truth itself as the higher-order or role property.

If we do so, then we can still be pluralists about the realizer properties of truth but without any threat of ambiguity. This means that by adopting functionalism, the alethic pluralist can explain universally applicable concepts like validity as well as can the alethic monist. Valid inference preserves truth. According to functionalism, truth is a higher-order functional property. Therefore, on the functionalist account, valid inference preserves that higher-order property. Of course, what makes this or that premise true (what realizes its truth) may be quite different from what constitutes or realizes the truth of the conclusion. But that is a question of deep metaphysics, not a formal, conceptual concern.

By acknowledging that truth is always and everywhere the property of playing the truth role, functionalism is not abandoning pluralism. Consider the case of a mental state like pain. According to functionalist accounts that take pain as a (higher-order) role property, there is indeed a sense in which the nature of pain is uniform across species. This is because, by and large, the pain role is uniform across species. But explaining the pain role does not explain what pain is in a more fundamental sense. It does not explain how that function is performed in a particular organism. For that, we must look to the details of the organism's neuronal structure: we must look for the lower-level property that realizes the pain role. In the same way, the functional role of truth does not explain how that role is filled in a particular discourse. For the underlying nature of truth, we must look to the details of the type of thought in question.

### 3 A (Brief) Case in Point: Juridical Truth

Let us consider a type of thought where the underlying nature of truth is plausibly other than correspondence with mind-independent facts. The domain in question is the law, or juridical truth. I emphasize that the

discussion will be extremely rough. My overall concern isn't with the nature of juridical truth itself. In this paper, juridical propositions are an example to aid in spelling out my main concern: the abstract structure of a functionalist theory.

When I write "juridical propositions," I have in mind what Ronald Dworkin calls "propositions of law" or "all the various statements and claims people make about what the law allows or prohibits or entitles them to have" (1986, 4). This includes both general claims, such as the proposition that segregation is illegal or that the law protects flag burning, and specific propositions, e.g., that Exxon must compensate Alaska for an oil spill. Propositions like this are surely capable of being true. We assume that in most cases, flag burning is or isn't protected under the law and that Exxon either is or is not required to pay compensation.

In a platitudinous sense, juridical propositions like these will be true when they correspond to the facts, tell it like it is, and so on. But in this case, it seems unwise to read the platitude as picking out a substantive metaphysical relation between propositions and mind-independent objects. Laws and legal entities are in some sense conventional; they are human constructs. But this hardly means that propositions of law are incapable of being true or false. It means only that it is unlikely that they are true in virtue of referential relations with mind-independent objects and properties. It is infinitely more plausible that, as Dworkin notes, "propositions of law are true or false (or neither) in virtue of other, more familiar kinds of propositions of which these propositions of law are (as we might put it) parasitic. These more familiar propositions furnish ... the 'grounds' of law" (1986, 4).

Dworkin's point is that legal propositions are *naturally* thought of as true because of their relation to other "grounding" propositions, not because they correspond to mind-independent objects called "laws" (or worse, "The Law"). Of course, just what is included within the grounds—what types of propositions, in other words, make propositions about the law true—is a matter of serious dispute. But at the very least, they include those propositions expressed in the Constitution, previous statutes, and past judicial decisions. It is these sorts of propositions that we think matter for whether it is true that a particular corporation is

required to pay compensation. Collectively, we might refer to them as the body of law.

This suggests a particular way of thinking about purely legal truth. Roughly, we think that a proposition of law is true when it coheres with its immediate grounds and with the grounds of propositions inferentially connected to it. In short, legal truth consists in coherence with the body of law.

A virtue of the functionalist theory is that it helps to explain this intuitive thought. Without the functionalist theory, the idea that legal truth consists in coherence with the body of law would seem to raise the ugly ambiguity problem. But with a functionalist story of truth in place, we can say that when it comes to propositions of law, the truth role is *realized* by a coherence relation of some sort. The fact that purely legal propositions concern conventional, mind-dependent matters does not mean that they are not really true or that they have a second-class status, but only that their truth is realized by a property different than what realizes truth about the physical world.

Of course, what it means for a proposition to cohere with the body of law would be (were this proposal worked out) a matter for substantive metaphysical inquiry. One question concerns the strength of the coherence relation involved. For example, for it to be true that flag burning is protected under the law, is it enough that this proposition cohere with the body of the law *now*? Or, weaker still, would it be enough that it was presently *believed* to cohere with the body of law? The latter certainly seems too weak. We need to allow room for mistakes about the law—for objectivity, in other words. One suggestion would be to understand juridical truth along the lines of what Wright calls “superassertibility.” A proposition is superassertible when it is “justified by some accessible state of information and will continue to be so justified no matter how that state of information is improved” (1999, 236). In other words, to be superassertible is to be durably justified without defeaters. Roughly, Wright’s view (1999, 228; see also this volume, chap. 32) is that superassertibility is a legitimate realizer of the “truth concept” for a class of propositions when those propositions are knowable and when evidence for or against such propositions is (in principle, at least) always accessible (1999, 236). These conditions seem to be met by legal truth. It is a

priori that legal truths are all knowable. It would be absurd to think that certain actions could be illegal without anyone even being capable of determining that they are. Second, the states of information that bear on the truth of juridical propositions, such as the empirical facts and the written laws themselves, are accessible to human investigation—at least in principle. Thus perhaps what makes a proposition of law true is that it durably or continually coheres with the body of the law. If so, then what is believed to be legal or illegal—even by a judge or the entire legal community—may not be. In short, juridical truth might turn out to be realized by “supercoherence” with the body of law, where a proposition can fail to have this property even if it coheres with the law in the short run, or coheres with judicial decisions that are later overturned.<sup>11</sup>

Much more would have to be said to make the case that juridical truth is realized by either coherence or supercoherence with a body of law. But I won’t be saying it here. I mention it simply as an illustration of what it would be for truth to be realized by something other than correspondence with mind-independent fact.

#### 4 A Note on Realization

Definitions like (FT) allow for multiple realization: in different contexts, different properties may occupy the functional role it describes. Yet (FT) describes that role in terms of the conjunction of all the alethic platitudes. If so, then one might wonder how any sort of epistemic property like supercoherence could play that role. For one may well think that one of our alethic platitudes is that every proposition is either true or false.<sup>12</sup> But it is a matter of serious dispute (see, e.g., Dworkin, 1986, 37 ff.) whether this platitude applies in the legal case. Further, one might think that if legal truth is a matter of coherence or supercoherence, then there may be juridical propositions that are neither completely true nor completely false. Perhaps there are cases where a proposition is neither supercoherent nor not supercoherent with a given body of law, because all possible evidence would fail to determine the matter. In any event, if such cases are possible and if juridical truth is realized by supercoherence, then not every juridical proposition need be determinately true or false. But how could this be if (FT) implies that whatever plays the truth role

must satisfy *all* alethic platitudes? The worry, in short, concerns how (FT) could be realized by such different types of propositions.

The solution is to see that a putative realizer property needn't satisfy every last one of the platitudes in *A* in order to count as realizing the truth role. After all, sophisticated functionalists in the philosophy of mind (those of the commonsense or analytical sort that we've been discussing) have long acknowledged that there might not be any single neural property that perfectly satisfies every folk-psychological platitude connected with pain (e.g., Lewis 1972). Indeed, it would be surprising if the world was so accommodating of our folk intuitions. It seems much more sensible, both in the alethic and psychological cases, to count properties that are nearly perfect in realizing the role in question as realizers of that role.

Recall that a functional role is a job, and (FT) is a job description. Suppose that we were wondering who was doing a certain job in a big company. It could turn out that no one person could fulfill management's written job description more than slightly. No one would then be doing that job. Or it could turn out that two or more people were each doing various aspects of the job, in which case, again, no one person would have that job. Or it could turn out that while no one could do every aspect of the job in that company, one person was in fact doing 90 percent of the job. In that case, no one was doing the job perfectly, but the job was getting done.

In the same fashion, we can allow that while there may be no perfect realizer in a discourse of the role marked out by (FT), there may be a property that is a near perfect realizer of that role in that discourse. In such cases, we can take whatever property uniquely fulfills the truth role near enough to be the property of truth in that discourse. If there is no property that comes close to realizing the role in the context, or if there are two or more properties that equally fulfill the role in that context, the discourse in question will remain unaletic.

Of course, what counts as "near enough" may be rather difficult to pin down precisely. Clearly, a property's fulfilling all but one of the platitudes in (FT) would seem to be near enough. Yet it seems reasonable to hold that the principles that comprise the alethic network may have different weight or importance within that network. Relatively central

principles like the equivalence schema or the idea that a proposition can be justified but not true will likely receive more weight than, e.g., the intuition that it is propositions that are true or false. Thus it may be that satisfying many of the heavily weighted principles will count more than satisfying all of the less heavily weighted ones. If this amendment is allowed, and there seems no reason why it shouldn't be, one can give truth a functional definition in terms of the total set of alethic platitudes but still allow that looser forms of thought can be true or false.

## 5 Objections

In discussing functionalism with other philosophers, I sometimes encounter puzzlement over how this theory differs from minimalism or deflationism about truth. I suspect that the main cause of this puzzlement is that the typical deflationist, like the functionalist, rejects the traditional correspondence and coherence accounts of truth. Further, the reasons given for this rejection are often similar, e.g., that the traditional theories fail to explain what truth is across the board (see, e.g., Horwich 1998, 1–2). Nonetheless, the lessons that minimalists draw from the failure of the traditional theories are quite different from those drawn by the functionalist. From the fact that the traditional theories can't explain how truth can have the same nature in every discourse, minimalists conclude that truth has no nature. This is often put by saying that the word "true" fails to express a genuine or substantive property of propositions. In contrast, the functionalist holds that truth does have a nature, and that "true" does express a substantive property.

This metaphysical difference between minimalists and functionalists about truth is like the difference between eliminativists and functionalists in the philosophy of mind. Eliminative materialism is the view that folk-psychological concepts like pain, belief, desire, etc., fail to pick out any actual physical properties or states. According to the eliminativist, there is, strictly speaking, no such thing as the nature of pain. The functionalist, on the other hand, holds that pain is a functional property, one that can be realized by distinct brain states in different species. The functionalist does not deny that pain has a nature; she holds that it can have more

than one. In the same way, an alethic functionalist holds that truth is a higher-order property that can be realized differently—or have a different nature—in different discourses.

Besides the metaphysical differences between deflationary theories and alethic functionalism, there is also a significant epistemological difference. Deflationists hold that truth is a philosophically unimportant concept. This is sometimes put by saying that we do not need to appeal to facts about truth in order to explain, e.g., knowledge, meaning, or any other philosophically troubling notion. Not so for functionalism. Just as understanding the underlying facts about how pain is realized is important for understanding a host of other psychological phenomena, so understanding the underlying facts about how truth is realized is essential (on the functionalist view, at least) for understanding a host of related philosophical concepts. Unlike the deflationist, the functionalist theory sees truth as having a job that impacts the entire philosophical economy.

Of course, there are differences between psychological and alethic functionalism. One might argue that these differences are greater than I've allowed, and as such, they undermine the case for calling alethic functionalism a type of "functionalism" at all. One objection of this sort is the following. In the case of mental concepts, analytic functionalists typically hold that it is a contingent a posteriori matter as to what occupies, e.g., the pain role. But surely it is an a priori matter whether some property plays the truth role, and if so, then perhaps this property must play that role in every discourse.<sup>13</sup> And this might be thought to undermine the claim of alethic functionalism that truth is multiply realizable.

In fact, there is not much difference between alethic and psychological functionalism on this matter. For according to alethic functionalism, truth in any discourse is the higher-order property of having a property that plays the truth role for *that* discourse. So realizers, as we noted above, are always *realizers for a discourse*. And as we saw in the case of legal discourse, what plays the truth role in any given discourse is determined by the nature of the discourse in question. Analogously, what plays the pain role for a given organism depends on that organism's material constitution. Given the way the organism is constituted, there will be one and only one property that best plays the pain role. And given



the way a discourse is constituted, there will be one and only one property that best plays the truth role. So what plays either sort of role is a contingent matter in that it is a contingent matter as to how any organism or any discourse is constituted. *Given that constitution*, what plays the role in question will do so necessarily. The difference between the two cases rests only in the differences in the subject matter. Figuring out what realizes the pain role in a particular organism is an a posteriori matter; figuring out what realizes the truth role in a particular discourse is (mostly) an a priori task. But these are epistemic facts; they do not impact the metaphysical question of multiple realizability at all.

A different type of worry is that functional definitions of alethic concepts like (FT) will be subject to what Michael Smith (1994) has called the “permutation” problem. According to Smith, the permutation problem arises for functional analyses of a type of concept when the platitudes used in the definitions are so tight-knit that they don’t connect the things to which the concept apply to things outside the network. When that happens, Smith suggests, we won’t have enough relational information left (after we’ve stripped the platitudes of the type of concepts we wish to define) to fix on any unique properties that we can identify as the realizers of the concepts. This means that there may be more than one property realizing every concept in the network. As a result, we lose the ability to distinguish the members of the network from each other.

As Smith acknowledges (1994, 54) one can never be sure that a given functional analysis is subject to the permutation problem unless one actually goes out and completes the analysis for each of the concepts in question. Fortunately, we can be confident that alethic functionalism is immune from the permutation problem without engaging in this task. Two points are relevant. First, there are plenty of platitudes connecting alethic concepts to other sorts of concepts. It is these platitudes that distinguish the various alethic concepts (and their realizers) from each other. Consider, for example, the differences elicited by the following simple principles: the fact that the butler’s prints are on the murder weapon can cause us to suspect that he is guilty of the crime, but no proposition, true or false, can cause anything; facts make propositions true, not vice versa; propositions can be doubted, facts cannot be; and so on.

Second, what determines the best realizer for the truth role isn't the alethic network alone. As I've repeatedly emphasized, part of what determines which property best realizes being true (or being a fact or being a proposition) in a discourse depends on a priori facts about that discourse. If, being committed physicalists, we were to take it as a priori that there could be no properties such as rightness or goodness, then the range of possible realizers for truth in the moral realm is significantly limited. Again, just as an organism's neural structure will determine what unique property best plays the pain role, so the constitution of a particular form of thought determines what unique property best plays the truth role.

The final objection I'll discuss also concerns functionalism's analysis of the concept of truth. In the philosophy of mind, Ramsey-Lewis definitions are frequently taken as reductions of our mental vocabulary to physical vocabulary (see Lewis 1972). The idea is that since functional definitions of mentality contain no mental terms (only variables), we can be said to have explained our mental concepts without circularity. One objection to alethic functionalism is that a moment's reflection indicates that the pool of *O* terms (terms that don't presuppose alethic concepts) in our alethic platitudes is too small to do the job required. There are bound to be concepts in (FT) that indirectly presuppose the concept of truth. As a result, (FT) cannot be an informative, noncircular analysis of truth.

The underlying assumption behind this objection is that functional definitions like (FT) can be informative only if they are completely reductive. But this is surely a fallacy. The functional method of definition supplies a way of giving the job descriptions for every property of the type in question all at once. But the fact that we can do so does not guarantee, either here or in the philosophy of mind, that we have entirely reduced the concepts in question to another set of concepts without remainder. And in the present case, our inability to reduce truth to more simple concepts is neither surprising nor avoidable. Arguably, truth is the most basic concept we possess; one cannot even get logic off the ground without some concept of truth. *Prima facie*, there is little reason to hope for *any* account of alethic concepts in terms that do not already presuppose them at some level of analysis or other. But to claim on this basis alone that a

functional specification of truth is uninformative is to assume without argument that the only informative analyses are those that are completely reductive. Yet clearly one can informatively and usefully explicate a concept without completely reducing it to more basic concepts. If we did have to live up to such high standards, we would be able to understand very few of our concepts indeed.

A more concessive answer to this objection is also possible for the functionalist. I have described alethic functionalism as (in part) a conceptual project, one that, roughly speaking, attempts to specify the meaning of alethic terms. Yet as Jackson has recently emphasized (1998, 143; see also Chalmers 1996, 61–64), there is more than one way to understand how a description of truth conditions like (FT) can “specify meaning.” The more traditional way, which I have thus far implicitly assumed, takes (FT) as *a priori* and necessary. On this understanding, (FT) says that *p*’s being true means that *p* has the higher-order property of having a property that plays the truth role for that discourse.<sup>14</sup> On the other hand, we can take “true” to “rigidly predicate” a particular property in every possible world in which it predicates anything at all.<sup>15</sup> If we do, we will take the right-hand side of (FT) as giving us the meaning of “true” in a very different sense, that is, as fixing the reference of “true.” But since “true” is a rigid predicator and the description on the right-hand side of (FT) is not, this would be similar to treating “is the man who taught Alexander” as what fixes the reference for “Aristotle.” And of course, Aristotle is only contingently the man who taught Alexander. On this alternative, (FT) is *a priori* but not necessary (Jackson 1998, 144). On this latter method, there is no barrier to employing other alethic concepts in our explanation of what “true” happens to ascribe to propositions. We take (FT) as fixing the reference of “true,” not as “defining” it in the traditional sense. Nonetheless, there is a sense of “meaning” in which (FT) is telling us the meaning of our concept of truth.

I see functionalism as engaged in the traditional project of meaning giving or conceptual explication without, as I’ve emphasized above, that project being necessarily reductive. But officially, the functionalist can remain neutral on the status of (FT). Taken either as a nonreductive conceptual explication or as fixing the reference, (FT) reveals important, *a priori* facts about the concept of truth.

## 6 Conclusion

A complete functionalist theory of truth has to meet at least three major demands. It must explain what it means to say that truth is a functional property; it must give a detailed account of how truth might be realized differently in distinct contexts; and it must explain the theory's impact on other important philosophical concepts, such as meaning, knowledge, or (what has come to be called) truth-aptness. In this paper I have been concerned with the first demand, having touched only briefly on the second. I have left the third aside altogether. Nonetheless, I think it is already clear that a complete functionalist theory would have significant advantages.

First, the functionalist theory of truth manages to satisfy both pluralist and monist intuitions. At one level, the functionalist theory is consistent with monism. In every discourse, the concept of truth is the concept of a particular higher-order property—the property of having the property that plays the truth role for that discourse. But at the level of deep metaphysics, alethic functionalism allows that this role may be realized by distinct properties that depend on the discourse in question. So to have the property of truth is to have a property that can, *by its very nature*, be realized in multiple ways.

Second, the functionalist theory explains the interrelatedness of the alethic concepts that has often made more traditional theorists about truth uncomfortable. A common reason for dismissing, e.g., “fact talk” among many deflationary-minded philosophers is that they are the mere shadows of propositions; that is, one cannot say what a fact is without invoking the notion of truth. But as just noted, this is what the functionalist would expect. In the philosophy of mind, it is common to think of belief and desire as so interrelated that the one cannot be explained without reference to the other. Yet this fact does nothing to show that such concepts, and the properties they are concepts of, are metaphysically suspect.

Third, the theory is ecumenical. Functionalism does not dictate in advance how truth will be realized in various discourses. To specify how truth is realized, we must look to the particulars of our thought. But once we do so, the functionalist theory may act as a neutral frame for a less

reductive picture of how our forms of thought relate to each other and to the world around us.

## Notes

Many people have helped me in writing this paper. The essential ideas were originally expressed in papers read at the 1999 Bled Epistemology Conference and at the Catholic University of Lublin, Poland. For conversations and correspondence I also thank William Alston, Terry Berthelot, Paul Bloomfield, Robert Barnard, Marian David, Hartry Field, Charles Fletcher, Terry Horgan, Frank Jackson, Peter Klein, Philip Pettit, Matjaz Potrc, Tadeusz Szubka, Robert Westmoreland, Michael Williams, Cory Wright, Crispin Wright, and audiences at the above places and at Connecticut College and the University of Southern Mississippi. Thanks also to the Master of St. Edmund's College, Cambridge, where, with the help of a grant from the University of Mississippi, I was able to complete a first draft of this paper.

1. See Quine (chap. 20), Horwich (chap. 24), Field (chap. 21), and Grover (chap. 22).
2. I refer here to the causal-relation theory of correspondence, espoused by Field (chap. 16) and Devitt (1984, see also chap. 25).
3. For objections of this sort to epistemic theories, see Lynch 1998, 107 ff., and Alston 1996. For a defense of the claim that epistemic theories work in some contexts, see the appendix to Wright 1999.
4. For a more detailed discussion of the problem of scope faced by correspondence, epistemic, and deflationist theories of truth, see Lynch 2000.
5. For further remarks on how my view differs from Horgan's and Wright's, see Lynch 1998, 129 ff.
6. Wright has emphasized (e.g., 1995, 215) that he too thinks that the concept of truth is univocal. In his contribution to this volume (chap. 32) Wright secures this univocality by distinguishing, much as I do below, the concept of truth and the underlying properties that realize truth. Nonetheless, as I note, Wright stops short of endorsing a functionalist account of alethic concepts.
7. I first began to suggest this idea in 1998, pp. 125 ff. Subsequently I learned that Philip Pettit had made the suggestion that truth could be understood functionally in his 1996 comments on Wright. Correspondence with Pettit confirms that our views on this matter are quite similar; I have also benefitted from Jackson's 1998 account of moral functionalism (although Jackson does not endorse the functionalist theory of truth).
8. For instance, in the philosophy of mind, one can understand functions computationally (Putnam), causally (Lewis, Armstrong), or biologically (Millikan, Dretske).
9. The analogy is Bob Barnard's (correspondence).

10. As I see it, there is no reason to settle which is which in advance. That is exactly the sort of information we would hope our theory would reveal. Thus, the platitude that a proposition is true *because* it corresponds to some fact may or may not mean that this correspondence with fact is the literal cause of the proposition's being true. The ordinary word "because" can also be used to mean "in virtue of," and  $x$  can be the case in virtue of  $y$  without  $x$  being causally related to  $y$ . Which relation "because" ends up picking out may depend, for all we know in advance, on the type of propositions and facts being considered.

11. Of course, lawyers and judges often disagree about what the law is; that is, they disagree about what is or should be included within the body of law. Often our decisions about what lies within or without the body of law depends in part on what we might call metalegal propositions, or propositions concerning a law's just or unjust nature, its accordance with "natural" law, its usefulness for society, and so on. Although the distinction cannot be an absolutely precise one, metalegal propositions are surely distinct from propositions of the law proper (or what we might call purely legal propositions). As such, we should expect that our opinions about metalegal matters (our opinions about what to include within the body of law) will be true or false in a different way than purely legal propositions, which the present proposal takes to be true in virtue of coherence with the body of law. See Walker's essay (chap. 6) for further discussion of what he calls a limited coherence theory of truth.

12. One reason that it is reasonable to suppose this is because an endorsement of the equivalence schema, together with a similar schema for falsity and the law of excluded middle, entails that a proposition is either true or false.

13. A related worry is this. According to recent "two-dimensional" semantics, terms can be understood as having both a primary and secondary intension (Chalmers 1995, 56 ff.; Jackson, 1998). A primary intension is a function from worlds considered as actual to a property, while a secondary intension is a function from worlds considered as counterfactual to a property. Now the primary and secondary intensions of "water" pick out different properties at a given world. When Putnam's Twin Earth world is considered as actual, "water" picks out XYZ, but when it is considered as counterfactual, "water" picks out  $H_2O$ . The worry is that any true functional term or concept must be similar to "water" in this respect. That is, it must have a primary intension that picks out different properties from what its secondary intension picks out in some worlds. But of course, "true" does not work like this. When I imagine that some other world is the actual world, I don't think that another property is being ascribed by "true"; indeed, the primary and secondary intension of the word "true" seem to be the same.

Like the problem above, this worry is unfounded. The key is to remember that on the functionalist theory I've presented here, truth is identical to the role property. For any discourse, truth is the higher-order property of having a property that plays the truth role in that discourse. Thus, like the higher-order property "is colored like the sky at this world" (where "this" refers to the world in question,

whether or not it is the actual world), the truth property will remain constant across worlds (discourses). But in both cases, it is obvious that different lower-level properties can realize the higher-order property. (I thank Philip Pettit for help with this point.)

14. If we were to take (FT) as giving the primary intension of the concept of truth (see last note) we would have to say that (FT) is necessarily true in *all possible worlds considered as actual* (Chalmers 1996, 63; also see Jackson 1997, chap. 2).

15. I take this term from Sydney Shoemaker 1975, 400.

## References

- Alston, W. P. 1996. *A Realist Conception of Truth*. Ithaca: Cornell University Press.
- Blackburn, S. 1998a. "Wittgenstein, Wright, Rorty, and Minimalism." *Mind* 107: 158–182.
- Blackburn, S. 1998b. *Ruling Passions*. Oxford: Oxford University Press.
- Blackburn, S., and Simmons, K. 1999. *Truth*. Oxford: Oxford University Press.
- Chalmers, D. 1996. *The Conscious Mind*. Oxford: Oxford University Press.
- Davidson, D. 1990. "The Structure and Content of Truth." *Journal of Philosophy* 87: 279–328.
- Davidson, D. 1967. "Truth and Meaning." *Synthese* 17: 304–323.
- Devitt, M. 1984. *Realism and Truth*. Princeton: Princeton University Press.
- Dworkin, R. 1986. *Law's Empire*. Cambridge: Harvard University Press.
- Field, H. 1972. "Tarski's Theory of Truth." *Journal of Philosophy* 69: 347–375.
- Field, H. 1995. "Disquotational Truth and Factually Defective Discourse." *Philosophical Review* 103, no. 3: 405–452.
- Gupta, A. 1993. "A Critique of Deflationism." Chap. 23 of this volume.
- Horgan, T. 1986. "Psychologism, Semantics, and Ontology." *Noûs* 20: 21–31.
- Horgan, T. 1991. "Metaphysical Realism and Psychologistic Semantics." *Erkenntnis* 34: 297–322.
- Horgan, T. 1996. "The Perils of Epistemic Reductionism." *Philosophy and Phenomenological Research* 56, no. 4: 891–897.
- Horwich, P. 1990. *Truth*. Cambridge: Basil Blackwell.
- Horwich, P. 1998. *Meaning*. Oxford: Oxford University Press.
- Jackson, F. 1998. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Oxford University Press.
- Kirkham, R. 1992. *Theories of Truth: A Critical Introduction*. Cambridge: MIT Press.
- Lewis, D. 1970. "How to Define Theoretical Terms." *Journal of Philosophy* 67: 427–446.

- Lewis, D. 1972. "Psychophysical and Theoretical Identifications." *Australasian Journal of Philosophy* 50, no. 3: 249–258.
- Lynch, M. 1997a. "Minimal Realism or Realistic Minimalism?" *Philosophical Quarterly* 47, no. 189: 512–518.
- Lynch, M. 1998. *Truth in Context*. Cambridge: MIT Press.
- Lynch, M. 2000. "Alethic Pluralism and the Functionalist Theory of Truth." *Acta Analytica* 15, no. 24: 195–214.
- Pettit, Philip. 1996. "Realism and Truth: A Comment on Crispin Wright's Truth and Objectivity." *Philosophy and Phenomenological Research* 56, no. 4: 883–889.
- Putnam, H. 1981. *Reason, Truth, and History*. Cambridge: Cambridge University Press.
- Putnam, H. 1994. "Sense, Nonsense, and the Senses: An Inquiry Into the Powers of the Human Mind." *Journal of Philosophy* 91, no. 9: 488–517.
- Sainsbury, M. 1996. "Crispin Wright: Truth and Objectivity." *Philosophy and Phenomenological Research* 56, no. 4: 899–904.
- Shoemaker, S. 1975. "Functionalism and Qualia." *Philosophical Studies* 27: 291–315.
- Smith, Michael. 1994. *The Moral Problem*. Cambridge: Blackwell.
- Wittgenstein, L. 1969. *On Certainty*. Cambridge: Blackwell.
- Wright, C. 1992. *Truth and Objectivity*. Cambridge: Harvard University Press.
- Wright, C. 1995. "Truth in Ethics." *Ratio* (new series) 8, no. 3: 210–226.
- Wright, C. 1996. "Response to Commentators." *Philosophy and Phenomenological Research* 56, no. 4: 911–941.
- Wright, C. 1998. "Comrades against Quietism." *Mind* 107: 184–203.
- Wright, C. 1999. "Truth: A Traditional Debate Reviewed." In Blackburn and Simmons 1999.





---

## Minimalism, Deflationism, Pragmatism, Pluralism

Crispin Wright

### 1 Minimalism and Deflationism: Overview

Deflationists have offered views about truth differing significantly in detail. But they characteristically maintain that as far as philosophy is concerned, there is nothing to say about truth that is not captured by a suitably generalized form of one (or both) of the following two schemata:

(ES) It is true that  $P$  iff  $P$ .

(DS) “ $P$ ” is true iff  $P$ .

And they maintain that this point in turn entails deflation—that the traditional metaphysical debates about truth, as well as more recent ones, are about nothing substantial.

It is worth noting that these are separable claims. Someone could allow that the two schemata—the Equivalence Schema, (ES), for propositions and the Disquotational Schema, (DS), for sentences—are each a priori correct<sup>1</sup> and (together) somehow fully encapsulate all proper uses of the truth predicate, without conceding that (it follows therefrom that) truth is somehow not a proper object of further philosophical enquiry, that no further metaphysical or semantic issues arise. Conversely, someone broadly in agreement with the antimetaphysical spirit of deflationism might hold that a correct characterization of the use of the truth predicate demands something more complicated than the two schemata.

The minimalist view about truth that I here defend rejects each of these deflationist claims, contending both that the two schemata are insufficient to capture all that should properly be reckoned as belonging to the concept of truth and that the antimetaphysical message of deflationism,

globally applied, represents a philosophical mistake.<sup>2</sup> Still, there are points of affinity between minimalism and deflationism. Minimalism agrees that, as far as the *conceptual* analysis of truth is concerned, matters should proceed by reference to set of basic a priori principles in which (ES) and (DS) are preeminent candidates for inclusion, and agrees too that aptitude for truth and falsity goes with surface assertoric content and is not the kind of deep property that, for instance, expressivist views about moral judgement standardly take it to be. However, minimalism rejects the idea that the analysis of the concept of truth exhausts the philosophy of truth: rather, even if the *concept* may be fully characterized by reference to certain basic a priori principles concerning it, the question of which *property* or *properties* of propositions, or sentences, realize the concept can still sensibly be raised for every discourse in which truth has application. Not that an answer to this question has necessarily to provide an identification of truth in the form “ $x$  is true iff  $x$  is  $F$ .” Minimalism only requires that each discourse that deals in truth-apt claims is associated with such a property whose character need not be fully determinable just from the list of basic principles serving to characterize the concept but which, relative to the discourse in question, serves as truth by dint of satisfying those principles. The fuller characterization of this property will depend on specific features of the particular discourse, and it will ultimately depend on these features whether or not the relevant truth property can be explicitly identified by, for instance, a biconditional of the type above.<sup>3</sup>

Minimalism thus incorporates a potential *pluralism* about truth, in the specific sense that what property serves as truth may vary from discourse to discourse. And it is this point which allows it to provide hospitality for the discussion of metaphysical—realist or antirealist—ideas that have fuelled those other traditional conceptions of truth that deflationists sought to undermine from the start. This potential pluralism is itself in opposition to the more traditional positions, insofar as they claim to uncover *the* universal nature of truth, something common to all truth-apt discourse. But it can still allow that some regions of discourse may be subject to a truth property congenial to broadly realist thinking about them, while in other regions the character of the truth property may be more congenial to antirealism.

All this may seem to suggest that the key difference between minimalism and deflationism resides in the fact that while the latter concedes the significance of the predicate “true,” and hence grants that there is a discussible *concept* of truth, it holds—in contrast to minimalism—that there is no *property* of truth: no property that all truths in a given area have in common.<sup>4</sup> This view of the matter would be encouraged by some of the literature in the field, but it is not the happiest way of putting the differences. For once the currency of a concept of truth is granted, it ought to be allowed that all truths have at least the following property in common: the property of falling under this concept.<sup>5</sup> No doubt this move may not illustrate the most natural or fruitful way of conceiving the relationship between concepts and their associated properties in general. But, for all that, it would be misleading to suggest that (most) deflationists would embrace the view that “Coal is black” and “Snow is white” have no more in common than do coal and snow.

The real distinction, then, between minimalism and deflationism in respect of the issue whether truth is a property is not that deflationism cannot consistently allow that it is, but rather that minimalism allows more: precisely, that the character of the property may not be transparent from the analysis of the concept. So in this respect there is a rough analogy with the relationship—to have recourse to a tired but useful example—between the concept of water and the property (that of being composed of H<sub>2</sub>O molecules, I suppose) that it denotes. Not that minimalism suggests that it should comparably be an a posteriori matter what property truth (locally) is. It will be a matter for further *conceptual* reflection—of a sort I will try to illustrate in the sequel—what (kind of) property best fulfills (locally) the role circumscribed by the concept. (That is why the water analogy is imperfect.)

This kind of substantial distinction between a concept, *F*, and the property it denotes, being *F*, is called for whenever we stand in need of some sort of general explanation of a characteristic of items that are *F* that cannot be elicited solely from materials directly implicated in those items’ falling under the concept in question.<sup>6</sup> To take a simple instance, suppose, to pursue the tired example, that the concept of water is a natural-kind concept after the fashion of Putnam’s well-known paradigm: that it is, e.g., given as the concept of that colorless, odorless,

tasteless liquid that is typically found in lakes and rivers, assuages thirst, and so on. If we allow that it makes good sense to ask *why* water typically presents with the surface features mentioned in its concept, we accept that there is a good explanatory question that cannot, obviously, be answered by appeal to water's falling under its concept, since we are asking for an explanation of the very features involved in its so doing. To allow the legitimacy of the question thus involves conceiving of whatever makes water what it is as distanced from the characteristics presented in its concept—as something that can potentially be invoked in explaining their habit of co-occurrence. But what makes water what it is is just its having the property of being water.

Now, it is plausible enough that there are no such explanations that might be given by appeal to the “thin” truth property that we envisaged the deflationist as admitting—the property of *falling under the concept of truth*—that we could not equally well give by appeal to the concept of truth itself. What the minimalist should claim, accordingly, in contradistinction to the deflationist, is that there *are* certain legitimate explanatory burdens that can be discharged only if we appeal to a property (or properties) of truth conceived in a more substantial sense of “property.”<sup>7</sup> And note that this claim can be true—in contrast with the situation of the kinds of explanation that might be given by appeal to the property of being water—even if truth, locally or globally, admits of no naturalistic (physicalistic) reduction. (It all depends on whether the things that need explaining are themselves so reducible.) As we shall see in due course, however, the minimalist's argument has no connection with the question of the feasibility of any such reduction.<sup>8</sup>

## 2 The Inflationary Argument

The inflationary argument is to the effect that the legitimacy of thinking of truth, in any particular discourse, as substantial in a fashion deflationism cannot accept, is already guaranteed by the very principles characterizing the concept of truth to which deflationism gives centre stage—at least when they are taken in conjunction with certain further uncontroversial principles. Thus minimalism does not just go beyond what deflationism allows but contends in addition that deflationism is incoherent: that, in

coupling the thesis that (ES) and/or (DS) yield(s) a complete account of truth with the contention that truth is a property only in the etiolated sense we have just reviewed, its proponents withdraw with one hand what they just tabled with the other.

We begin with the observation that truth-apt contents, or sentences expressing such contents, demand a distinction between circumstances under which asserting them is warranted and those under which it is not. And competent thought and talk requires an ability to tell the difference: I need to be able to tell which assertions I am warranted in making in a given state of information and which I am not. So if I am warranted in asserting *P*, that fact will be recognizable to me, and I will thereby be warranted in claiming that I am so warranted. Conversely, if I am warranted in thinking that the assertion of *P* is warranted, I will be beyond relevant—that is, epistemic—reproach if I go on to assert it. But that is to say that I will be warranted in doing so. We accordingly obtain:

There is warrant for thinking that [it is warrantably assertible that *P*] iff there is warrant for thinking that [*P*].

Given the Equivalence Schema, this will in turn yield:

There is warrant for thinking that [it is warrantably assertible that *P*] iff there is warrant for thinking that [it is true that *P*].

And now, since warranted assertibility is, in a perfectly trivial sense, a normative property—a property possession or lack of which determines which assertions are acceptable and which are not—it follows that truth is too. For by the above equivalence, to be warranted in thinking that *P* is true has exactly the same normative payload as being warranted in thinking that it is warrantably assertible. Moreover, our finding is that truth, as characterized by the schemata, and warranted assertibility *coincide in positive normative force*.

That is hardly a startling finding. But the relevant point is not the result itself but its provenance: that truth's being normative in the fashion noted is not merely plausible anyway but is a consequence of what ought to be uncontroversial considerations about the concept of assertibility and a central tenet of deflationism: the conceptual necessity of the Equivalence Schema. However, given only the further assumption that any

$P$  apt for truth has a significant negation that is likewise apt for truth, the Equivalence Schema will also entail any instance of the following *Negation Equivalence*:<sup>9</sup>

(NE) It is true that [not- $P$ ] iff it is not true that [ $P$ ].

And this shows that, coincident in positive normative force though they may be, we cannot in general *identify* truth and warrant. For most propositions about most subject matters allow of neutral states of information: states of information in which there are neither warrants for asserting  $P$  nor for asserting its negation. In any such case, an invalid schema results if we substitute “is warrantably assertible” for “is true” in (NE). More specifically, if the propositions that make up the substitution class for  $P$  allow in principle of neutral states of information, the following conditional is *not* valid:

It is warrantably assertible that [not- $P$ ] if it is not warrantably assertible that [ $P$ ].

Thus, we can already conclude from (NE), and hence from (ES), that truth and warranted assertibility, even if coinciding in positive normative force, are *potentially divergent in extension*.<sup>10</sup>

It is an immediate consequence of this observation that for any assertoric practice that allows the definition, on the contents of the moves it permits, of a truth property satisfying (ES)—that is, for any assertoric practice whatever—there must be a further kind of distinction between circumstances in which making these moves is in good standing and circumstances in which it is not—a distinction that need not coincide with the distinction between circumstances in which such a move can warrantably be made and those in which it cannot. The concept of truth as characterized by (ES) precisely calls for a norm—a way an assertion may be in good standing—which warrant is essentially warrant to suppose satisfied but which, because of the point about potential extensional divergence, may nevertheless not be satisfied when an assertion is warranted (or may be satisfied when it is not). And a fully intelligent participation in such practices will involve grasp that they essentially involve submission to a standard the meeting of which need not just be a matter of possessing warrants for the claim that it is met.

Minimalism now claims that these facts about assertoric practices stand in need of explanation. In particular, it maintains that it needs to be explained what this further norm of correctness amounts to in such a way that it becomes clear how it and warranted assertibility, although potentially divergent in extension, coincide in normative force: how it can be that warrant is essentially warrant to think that this other norm is satisfied when there is no guarantee that they are always *co-satisfied*. And such an explanation, it is contended, while it will have to do much more than this, must at least begin by finding something for the truth of a proposition to consist in, a property that it can intelligibly have although there may currently be no reason to suppose that it has it, or may intelligibly lack even though there is reason to think that it has it. Warrant can then be required to be whatever gives a (defeasible) reason to think that a proposition has that property.

The deflationist account of truth would appear, however, to have no resources to give such an explanation. For all we can elicit from the Equivalence Schema is the *problem*. The point of the inflationary argument is precisely that the basic principles on which deflationism builds its account spawn the concept of a norm—a way a proposition can be in good or bad standing, as I put it a moment ago—that contrasts with its current evidential status. But these principles keep silence when the question is raised, What does the satisfaction or nonsatisfaction of this new norm consist in, and how can it fail to be a substantial property?

So at any rate the inflationary argument contends. But the deflationist is likely to believe that she has a good response. “There is no silence on the point,” she will reply. “On the contrary, my theory is very explicit about what the satisfaction of your ‘norm’ consists in. The proposition that snow is white satisfies it just if snow is white; the proposition that grass is green satisfies it just if grass is green, the proposition that there is no life on Mars satisfies it just if there is no life on Mars...” However, this response is, of course, to no avail unless we *already* understand the difference between the proposition, e.g., that there is no life on Mars and the proposition that that proposition is warranted. And clearly this distinction cannot be recovered from any contrast between the circumstances under which the two propositions are *respectively* warranted, since—as in effect noted right at the start of the argument—there is none.



The difference between them resides, rather, precisely in a difference in correctness conditions of another sort (whisper: *truth* conditions): in order to understand the contrast between the two propositions, I precisely have to understand that the former is in principle hostage to a kind of failure that can occur even when it is warranted, and that will not then affect the latter. So the debate is rapidly brought back to the point before the deflationist made her putative “good response,” with the minimalist charging her to explain (i) how the relevant contrast can so much as exist unless there is something substantial in which such failure—or more happily, success—consists, and (ii) how a *grasp* of the contrast can anywhere be possible unless we are familiar with a (perhaps local) property that behaves as the concept characterized by the basic principles demands.

The kind of move we just envisaged a deflationist making is, of course, pure deflationist stock-in-trade. Supporters of deflationism characteristically view the whole debate as turning on whether it can be shown that all legitimate uses of the word “true” can somehow be explained on the basis of the Equivalence Schema (and/or the Disquotational Schema) together with a repertoire of contexts free of “true” and its cognates, and they put all their effort and (often considerable) ingenuity into the attempt to show that these uses can be so explained.<sup>11</sup> But success in this project is entirely beside the point if the contents of the relevant “true”-free contexts, to which deflationists simply help themselves, cannot be explicated by construing them merely as subject to norms of assertibility but demand an additional truthlike constraint. Deducing some aspect of our use of the predicate “true” by appeal just to the Equivalence Schema and certain “true”-free contexts cannot just be assumed to have reductive significance without further ado. The initial position in the debate is one in which nothing yet stands against the opposed thought that, instead of reading the Equivalence Schema from left to right, as if to eliminate the truth property, we should read it from right to left, as highlighting the fact that, implicit in any content in the range of “*P*,” there is already a tacit invocation of the norm of truth. Deflationism needs to get to grips with this reading: to make a case that no implicit prior grasp of the concept of truth, nor implicit reference to a property that the concept denotes, lurks buried in the materials to which its “explanations” appeal. The thrust of

the inflationary argument is that no such convincing case can be made—that whether or not we can somehow eliminate or otherwise “deflate” the *word*, a corresponding property, and its contrast with assertibility, is part and parcel of assertoric content itself.<sup>12</sup>

### 3 Pragmatism and Pluralism (I): Peirce and Putnam

Let me now be a little more explicit about how minimalism opens up prospects for a pluralistic conception of truth. Above, I spoke approvingly of the idea, of which the deflationist proposals can be seen as one example, that as far as the conceptual analysis of truth is concerned, matters should proceed by reference to a set of basic a priori principles variously configuring or bearing on the concept. Many philosophers, from Frege to Davidson, have, of course, doubted whether truth allows of any illuminating philosophical analysis. But their skepticism has been driven largely by the traditional notion that success in this project would have to consist in the provision of a satisfactory necessary-and-sufficient-conditions analysis of the concept, and there is clearly some scope for relaxation of that model. After all, such a necessary-and-sufficient-conditions analysis, even if it could be provided, would only culminate in one particular a priori—presumably, conceptually necessary—claim. Why should not other such claims—even if not biconditional or identity claims—provide illumination of essentially the same kind? To be sure, if one wants a priori conceptual clarity about what truth—or beauty, or goodness, etc.—is, then the natural target is an identity (or a biconditional). But perhaps the sought-for reflective illumination can be equally well—if less directly—provided by the assembly of a body of conceptual truths that, without providing any reductive account, nevertheless collectively constrain and locate the target concept and sufficiently characterize some of its relations with other concepts and its role and purposes.

What should such principles be for the case of truth? The method here should be initially to compile a list, including anything that chimes with ordinary a priori thinking about truth—what I shall call a *platitude*—and later to scrutinize more rigorously for deductive articulation and for whether candidates do indeed have the right kind of conceptual plausibility. So we might begin by including, for instance,

- the transparency of truth—that to assert is to present as true and, more generally, that any attitude to a proposition is an attitude to its truth—that to believe, doubt, or fear, for example, that *P* is to believe, doubt, or fear that *P* is true. (*Transparency*)
- the epistemic opacity of truth—incorporating a variety of weaker and stronger principles: that a thinker may be so situated that a particular truth is beyond her ken, that some truths may never be known, that some truths may be unknowable in principle, etc.<sup>13</sup> (*Opacity*)
- the conservation of truth-aptitude under embedding: aptitude for truth is preserved under a variety of operations—in particular, truth-apt propositions have negations, conjunctions, disjunctions, etc., which are likewise truth-apt. (*Embedding*)
- the Correspondence Platitude—for a proposition to be true is for it to correspond to reality, accurately reflect how matters stand, “tell it like it is,” etc. (*Correspondence*)
- the contrast of truth with justification—a proposition may be true without being justified, and vice versa. (*Contrast*)
- the timelessness of truth—if a proposition is ever true, then it always is, so that whatever may, at any particular time, be truly asserted may—perhaps by appropriate transformations of mood, or tense—be truly asserted at any time. (*Stability*)
- that truth is absolute—there is, strictly, no such thing as a proposition’s being more or less true; propositions are completely true if true at all. (*Absoluteness*)

The list might be enlarged, and some of these principles may anyway seem controversial. Moreover, it can be argued that the Equivalence Schema underlies not merely the first of the platitudes listed—Transparency—but the Correspondence Platitude<sup>14</sup> and, as we have seen in discussion of deflationism, the Contrast Platitude as well.

There’s much to be said about this general approach to conceptual analysis, and many hard and interesting questions arise, not least, of course, about the epistemological provenance of the requisite basic platitudes. But such questions arise on *any* conception of philosophical analysis, which must always take for granted our ability to recognize basic truths holding a priori of concepts in which we are interested.

Let us call an account based on the accumulation and theoretical organization of a set of such platitudes concerning a particular concept an *analytical theory* of the concept in question.<sup>15</sup> Then the provision of

an analytical theory of truth in particular opens up possibilities for a principled pluralism in the following specific way: in different regions of thought and discourse, the theory may hold good a priori of—may be satisfied by—different properties. If this is so, then always provided the network of platitudes integrated into the theory is sufficiently comprehensive, we should not scruple to say that truth may consist in different things in different such areas: in the possession of one property in one area, and in that of a different property in another. For there will be nothing in the idea of truth that is not accommodated by the analytical theory, and thus no more to a concept's presenting a truth property than its validating the ingredient platitudes. In brief, the *unity* in the concept of truth will be supplied by the analytical theory, and the *pluralism* will be underwritten by the fact that the principles composing that theory admit of *variable collective realization*.

An illuminating case study for these ideas is provided by pragmatist conceptions of truth. In a very famous passage, C. S. Peirce writes,

Different minds may set out with the most antagonistic views, but the progress of investigations carries them by a force outside themselves to one and the same conclusion. This activity of thought by which we are carried, not where we wish but to a fore-ordained goal, is like the operation of destiny. No modification of the point of view taken, no selection of other facts to study, no natural bent of mind even, can enable a man to escape the predestinate opinion. This great law is embodied in the conception of truth and reality. *The opinion which is fated to be ultimately agreed by all who investigate is what we mean by the truth*, and the object represented in this opinion is the real. [My italics]<sup>16</sup>

Here Peirce seemingly believes in a predestined march towards a stable scientific consensus among “all who investigate,” but the received understanding of the “Peircean” view, whether historically faithful or not, has come to be, rather, that the true propositions are those on which investigators *would* agree if—which may well not be so—it were possible to pursue enquiry to some kind of ideal limit; that

*P* is true if and only if, were epistemically ideal conditions to obtain, *P* would be believed by anyone who investigated it.

An equally famous passage in Hilary Putnam's *Reason, Truth, and History* has regularly been interpreted as advancing the same proposal. Having rejected the identification of truth with what he calls rational acceptability, Putnam there suggests that

truth is an *idealisation* of rational acceptability. We speak as if there were such things as epistemically ideal conditions, and we call a statement “true” if it would be justified under such conditions.<sup>17</sup>

He explains that, as he intends the notion, “epistemically ideal conditions” are an idealization in the same way that frictionlessness is: they are conditions that we cannot actually attain, nor—he adds, interestingly—can we “even be absolutely certain that we have come sufficiently close to them.” He is explicit that he is not “trying to give a formal *definition* of truth, but an informal elucidation of the notion.” And he goes on to say that

the two key ideas of the idealisation theory of truth are (i) that truth is independent of justification here and now, but not independent of *all* justification. To claim a statement is true is to claim it could be justified. (ii) Truth is expected to be stable or “convergent.”<sup>18</sup>

Putnam has, of course, since officially moved a long way from these ideas.<sup>19</sup> But this is the nearest that he ever came to explicitly endorsing the Peircean conception, and it is clear that his words left considerable latitude for interpretation. In particular, there was no unmistakable suggestion of a key feature of the Peircean proposal: that some *single* set of “epistemically ideal conditions” would be apt for the appraisal of any statement whatever.

Putnam himself subsequently returned to emphasize that point. In the Preface to *Realism with a Human Face* he again endorsed the idea that to claim of any statement that it is true is, roughly, to claim that it could be justified were epistemic conditions good enough.<sup>20</sup> And he goes on to allow that “one can express this by saying that a true statement is one that could be justified were epistemic conditions ideal.” But then he proceeds immediately to repudiate the idea

that we can sensibly imagine conditions which are *simultaneously ideal* for the ascertainment of any truth whatsoever, or simultaneously ideal for answering any question whatsoever. I have never thought such a thing, and I was, indeed, so far from ever thinking such a thing that it never occurred to me even to warn against this misunderstanding. . . . I do not by any means *ever* mean to use the notion of an “ideal epistemic situation” in this fantastic (or utopian) Peircean sense.<sup>21</sup>

Rather, the notion of ideal epistemic circumstances stands in need of specialization to the subject matter under consideration:

If I say “there is a chair in my study,” an ideal epistemic situation would be to be in my study, with the lights on or with daylight streaming through the window, with nothing wrong with my eyesight, with an unconfused mind, without having taken drugs or being subjected to hypnosis, and so forth, and to look and see if there is a chair there.

Indeed, we might as well drop the metaphor of idealisation altogether. Rather, “there are *better and worse* epistemic situations *with respect to particular statements*. What I just described is a very good epistemic situation with respect to the statement ‘there is a chair in my study’.”<sup>22</sup>

These remarks might invite the following regimentation. Let us, for any proposition *P*, call the following the *Peircean biconditional* for *P*:

*P* is true if and only if, were *P* appraised under conditions *U*, *P* would be believed,

where *U* are conditions under which thinkers have achieved some informationally comprehensive ideal limit of rational-empirical enquiry. And let us call the following the corresponding *Putnamian biconditional* for *P*:

*P* is true if and only if, were *P* appraised under topic-specifically sufficiently good conditions, *P* would be believed.

Then we now have two contrasting pragmatist conceptions of truth to consider. And the question is, Do they—either of them—meet the standard set by our proposed minimalism: do they realize the relevant constitutive platitudes?<sup>23</sup>

There is an interesting difficulty about an affirmative answer. Putnam imposed what he termed a convergence requirement on his conception of truth—that there be no statement such that both it and its negation are assertible under epistemically ideal (topic-specifically sufficiently good) conditions.<sup>24</sup> This is to be distinguished, of course, from any requirement of *completeness*. The requirement of completeness would be that, for each statement, *either* it *or* its negation must be justified under such circumstances. There seems no good reason to impose any such completeness requirement—no particular reason why all questions that are empirical in content should become decidable under Peirce’s or Putnam’s respective ideal conditions. Indeed, to take seriously the indeterminacies postulated by contemporary physical theory is to consider that there is

reason to the contrary. We should expect that a pragmatist would want to suspend the Principle of Bivalence for statements that would find themselves in limbo under epistemically ideal, or topic-specifically sufficiently good, conditions in this way, and ought consequently, one would imagine, to want to suspend it in any case, failing an assurance that no statements are actually in that situation.

So what is the promised difficulty? That there is, apparently, a simple inconsistency within the triad uniting either of our pragmatist biconditionals with the claim that the notion of truth it concerns complies with the minimal platitudes, and the admission that certain statements may remain undecidable under epistemically ideal, or topic-specifically sufficiently good, circumstances, neither they nor their negations being justified. For, as we have seen, the minimal platitudes impose the standard Negation Equivalence:

(NE) It is true that [not- $P$ ] iff it is not true that [ $P$ ].

And to allow that, even under epistemically ideal or topic-specifically sufficiently good circumstances, we might yet be in a state of information that provided warrant neither for  $P$  nor for its negation would force us to reject the right-to-left ingredient in (NE) when “true” is interpreted in accordance with either pragmatist biconditional. In other words, it seems that epistemically ideal or topic-specifically sufficiently good circumstances cannot be neutral both on a statement and its negation if the Equivalence Schema is in force over all assertoric contents, if every assertoric content has a negation that is an assertoric content, and if truth is Peircean or Putnamian.

Simple though this train of thought is, it provides, on the face of it, a devastating blow to both pragmatist proposals. Leave on one side the obvious difficulties occasioned by the undecidability of mathematical examples like, say, the generalized continuum hypothesis. Surely, it should not be true a priori even of empirical statements in general that each would be decidable—confirmable or disconfirmable—under epistemically ideal or topic-specifically sufficiently good circumstances. But the relevant minimal platitudes, for their part, presumably hold true a priori. So *if* either pragmatist proposal were a priori correct—as it has to be if it is correct at all—it would have to be a priori that if a statement failed to

be justified under epistemically ideal or topic-specifically sufficiently good circumstances, its negation would be justified instead—just the thing, it seems, that cannot be *a priori*. Invited conclusion: such proposals incorporate mistaken *a priori* claims about the concept of truth, and the properties they present are hence unfitted to serve as realizers of that concept.

Indeed, the point is more general: a simple extension of the argument seems to tell not just against the two tabled pragmatist proposals but against any attempt to represent truth as essentially evidentially constrained. Someone in sympathy with Dummettian antirealism, for instance, may content herself with a one-way Principle of Epistemic Constraint,

(EC) If *P* is true, then evidence is available that it is so.

Yet still be posed an embarrassment by the argument. For if no evidence is available that *P*, then, contraposing on (EC), she ought to allow that it is not the case that *P* is true, whence, by the Negation Equivalence, its negation must count as true. So in the presence even of a one-way epistemic constraint, the unattainability of evidence for a statement is bound, it appears, to confer truth on, and hence, via (EC), to ensure the availability of evidential support for, its negation—contrary to what, someone might very well think, the antirealist could and should admit, namely, that some statements may be such that no evidence bearing upon them is available either way, even under idealized conditions of investigation. (Indeed, how do we explain the semantical antirealist's characteristic refusal to allow the unrestricted validity of the Principle of Bivalence unless it is based on precisely that admission, coupled with the insistence that truth is evidentially constrained?)

What room does such an antirealist have for maneuver here? We can take it that, unless she decides to off-load the notion of truth entirely, there is no denying the Equivalence Schema. Maybe trouble might somehow be found for the move from that to the Negation Equivalence. But the prospects do not look bright.<sup>25</sup> What is needed, rather, is a way to reconcile the Negation Equivalence with an insistence that truth is evidentially constrained and the admission that not every issue can be guaranteed to be decidable, even in principle. But is there any scope for such a reconciliation?



Yes, there is. There can be no denying that the Negation Equivalence commits someone who endorses (EC) to allowing (A):

- (A) If no evidence is available for  $P$ , then evidence is available for its negation.

And, of course, it's extremely easy to hear this as tantamount to the admission that evidence is in principle available either for affirming  $P$  or denying it. But there is a suppressed premise in this turn of thought: the premise (B), an instance of the law of excluded middle:

- (B) Either evidence is available for  $P$  or it is not.

Classically, of course, the conditional (A) is an equivalent of the disjunction (C):

- (C) Either evidence is available for  $P$ , or evidence is available for its negation.

But the proof of the equivalence depends on the instance of the law of excluded middle, (B). If we may not assume that evidence either is or is not available for an arbitrary statement, then the convertibility of lack of evidence for a particular statement into evidence for its negation, demanded by the Negation Equivalence when truth is evidentially constrained, need not impose (C), and so need not be in contradiction with the a priori unwarrantability of the claim that the scales of (in principle) available evidence must tilt, sooner or later, one way or the other, between each statement and its negation.

This is a substantial result. It teaches us, in effect, that in order to sustain the claim of our two pragmatist proposals—and indeed any broadly Dummettian antirealist proposal—to offer defensible conceptions of truth, the associated package must include revisions of classical logic *of a broadly intuitionistic sort*. For otherwise there is no possibility of modeling the minimal platitudes consistently with a proper recognition that decidability is often not guaranteed even under ideal—Peircean or Putnamian—conditions.

Provided this way of surmounting the difficulty is accepted, our two pragmatist proposals remain in the field as offering two possible ways in which a property satisfying the minimal constraints on truth may be constructed out of assertibility by idealization. Of course, we have only

considered just one problem, so the proposals' claim to succeed in that regard would need more detailed review. But I shelve consideration of that review to turn to another serious and independent form of difficulty confronting each of them, a difficulty that, I contend, should force a pragmatist-inclined philosopher to look for a subtly different kind of conception of truth.

Here is a generalization (and, in one respect, simplification—see note 28) of an objection advanced by Alvin Plantinga specifically against the Peircean proposal.<sup>26</sup> Assume any purported account—or indeed any “informal elucidation”—of truth of the form (o):

(o) It is true that  $P \leftrightarrow (Q \Boxrightarrow Z(P))$

where  $Q$  expresses a general epistemic idealization,  $Z( )$  is any condition on propositions—for instance, being judged to be true by the ideally rational and informed thinkers whose existence is hypothesized by  $Q$ , or cohering with the maximally coherent set of beliefs whose existence is hypothesized by  $Q$ , etc.—and “ $\Boxrightarrow$ ” expresses the subjunctive conditional. Since (o) is purportedly a correct elucidation of a concept, it presumably holds as a matter of conceptual necessity. Thus:

(i) Necessarily (It is true that  $P \leftrightarrow (Q \Boxrightarrow Z(P))$ )

Now suppose that (ii):

(ii) Possibly ( $Q \ \& \ \text{Not } Z(Q)$ )

Then, by logic and the Equivalence Schema, (iii):

(iii) Possibly (It is true that  $Q \ \& \ (Q \ \& \ \text{Not } Z(Q))$ )

But (iii) contradicts (i), with “ $Q$ ” taken for “ $P$ ,”<sup>27</sup> which therefore entails

(iv) Not possibly ( $Q \ \& \ \text{Not } Z(Q)$ )

So

(v) Necessarily ( $Q \rightarrow Z(Q)$ )

A necessarily true conditional ought to be sufficient for the corresponding subjunctive, so:

(vi)  $Q \Boxrightarrow Z(Q)$

So, from (i):

(vii) It is true that  $Q$

So by the Equivalence Schema again:

(viii)  $Q$

The upshot is, it seems, that anyone proposing an account of truth of the shape typified by (o) must accept that the idealization  $Q$  *already obtains*. Thus the Peircean must accept that conditions are already “epistemically ideal,” (and a coherence theorist must accept that there already is a controlled, comprehensive, and coherent set of beliefs.)<sup>28</sup> Obviously, this is unacceptable. And it is not clear how the Peircean can respond.

However, just here is where there may seem additional point to the more modest Putnamian proposal. For the key to the proof above is the license, granted by the Peircean conception of truth in particular, to assume that the conditions that are ideal for the appraisal of the proposition  $U$  are the very conditions depicted by that proposition—it is this assumption that sanctions the substitution of “ $Q$ ” for “ $P$ ” in (o). Suppose instead that, with erstwhile Putnam, the pragmatist drops the idea of such a comprehensive set of epistemically ideal conditions and that (o) gives way to a range of Putnamian biconditionals:

(o') It is true that  $P \leftrightarrow (Q_P \sqsupset Z(P))$

Here  $Q_P$  is the hypothesis that conditions are sufficiently good for the appraisal specifically of  $P$ . We can advance as before to:

(iii') Possibly (It is true that  $Q_P \ \& \ (Q_P \ \& \ \text{Not } Z(Q_P))$ )

But nothing harmful need follow unless one of our Putnamian biconditionals is:

It is true that  $Q_P \leftrightarrow (Q_P \sqsupset Z(Q_P))$

which will be available only if conditions  $Q_P$  are topic-specifically sufficiently good not merely for the appraisal of  $P$  but also for the appraisal of the proposition  $Q_P$  itself, that is, if  $Q_P = Q_{Q_P}$ . And why should that be so?

Well, but the question should be, Is it certain such an identity is *never* realized? Consider Putnam's own example: a sufficiently good epistemic situation for appraisal specifically of “There is a chair in my study.” That would be, he said, to be in my study, with the lights on or with daylight

streaming through the window, with nothing wrong with my eyesight, with an unconfused mind, without having taken drugs or being subjected to hypnosis, and so forth. But wouldn't these conditions likewise be sufficiently good conditions in which to appraise the claim that I was indeed in my study, with the lights on or with daylight streaming through the window, with nothing wrong with my eyesight, with an unconfused mind, without having taken drugs or being subjected to hypnosis, and so forth? Maybe not—maybe there is some condition whose addition to the list would not improve my epistemic situation with respect to “There is a chair in my study” but without which I would not be best placed to assess the complex proposition just stated. But even if so in the particular example, must that *always* be so? Unless we can see our way to justifying an affirmative answer, there can be no assurance that Plantinga's problem can be resolved by a fallback to Putnamian biconditionals.

In fact it is clear that the most basic problem with the Peircean biconditional cannot be resolved by this fallback. Plantinga made a difficulty by taking  $Q$  for  $P$  in (o). But suppose instead we take “ $Q$  will never obtain,” thus obtaining:

$Q \text{ will never obtain} \leftrightarrow (Q \Box \rightarrow Z(Q \text{ will never obtain}))$

Then if the right hand side is interpreted as in the Peircean biconditional, we have a claim to the effect that conditions will always be less than epistemically ideal just in case thinkers who considered the matter under epistemically ideal conditions would suppose so. This is obviously unacceptable. And it is an illustration of a very general point: that no categorical claim  $P$  can be a priori (or necessarily) equivalent to a subjunctive conditional of a certain type—roughly, one whose antecedent hypothesizes conditions under which a manifestation, depicted by the consequent, of the status of  $P$  takes place—unless it is likewise a priori (or necessary) that the realization of the antecedent of the latter would not impinge on the actual truth value of the categorical claim. More specifically, it cannot be a priori—or necessary—that

It is true that  $P \leftrightarrow$  were conditions  $C$  to obtain, such and such an indicator  $M$  of  $P$ 's status would also obtain

unless it is a priori (or necessary) that the obtaining of  $C$  would not bring about any change in the actual truth value of  $P$ . For suppose that it is

true that  $P$ , but that were conditions  $C$  to obtain, it would cease to be so: would  $M$  then obtain? Yes. For by hypothesis,  $P$  is actually true. So the biconditional demands that  $M$  would obtain if  $C$  did. So not- $P$  would hold alongside conditions  $C$  and  $M$ . But in that case  $M$  would not be an indicator of  $P$ 's status in those circumstances after all. In particular, if  $M$  consists in the believing that  $P$ , suitably placed thinkers, then the effect will be that their beliefs will be *in error* under conditions  $C$ —exactly what the pragmatist proposal was meant to exclude.

This point—or anyway the general thought, epitomized in the phrase, “The Conditional Fallacy,” that subjunctive conditional analyses are almost always unstable—is nowadays very familiar from the literature on dispositions and response-dependence.<sup>29</sup> What is clear for our present purpose is that it is no less a problem for Putnamian biconditionals than for Peircean ones. That is, unless it is given a priori that the implementation of conditions  $Q_P$  would not impinge on the circumstances actually conferring its truth value on  $P$ , it cannot be supposed to hold purely in virtue of the concepts involved that

It is true that  $P \leftrightarrow (Q_P \squarerightarrow P \text{ would be believed})$

except at the cost of allowing that even under  $Q_P$  circumstances,  $P$  might be believed when false. And again, this is just to surrender the idea that belief under ideal circumstances is guaranteed to line up with the facts: the cardinal tenet of this kind of pragmatism.<sup>30</sup>

#### 4 Pragmatism and Pluralism (II): Superassertibility

The ur-thought behind any pragmatist conception of truth is that the notion should be grounded in ordinary human practices of assessment and epistemic values. So some form of idealized assertibility is the most natural concrete interpretation of the idea. But I think that the Peircean and Putnamian conceptions idealized assertibility in the wrong direction. Warranted assertibility is assertibility relative to a state of information. So it can seem as if there is only one direction for a truthlike idealization of assertibility to assume: to wit, we have somehow to idealize the state of information involved, as both the Peircean and Putnamian proposals do in their different ways. But there is another way. Rather than ask

whether a statement would be justified at the limit of ideal empirical investigation or under topic-specifically sufficiently good circumstances, whatever they are, we can ask whether an ordinary carefully controlled investigation, in advance of attaining any mythical or more practical limit, would justify the statement, and whether, once justified, that statement would continue to be so no matter how much further information were accumulated.

More carefully, another property constructible out of assertibility that is both absolute and, so it is plausible to think, may not be lost—Putnam's two desiderata—is the property of being justified by some (in principle accessible) state of information and then *remaining* justified no matter how that state of information might be enlarged upon or improved. Like Peircean truth, the characterization of this property presupposes that we understand what it is for one state of information to enlarge upon or otherwise improve another. But it does not presuppose that we grasp the idea of a limit to such improvement—a state of information that is itself beyond all improvement—or even have any general conception of what it would be for the topic-specific epistemic circumstances to be unimprovable. So this characterization need not confront questions about the intelligibility and coherence of the idea of the Peircean limit, nor need it confront the question of how appraisal under merely topic-specifically sufficiently good conditions can guarantee the stability of a verdict, and thereby the stability of Putnamian truth.

Elsewhere I have called the property just prefigured *superassertibility*.<sup>31</sup> A statement is superassertible, then, if and only if it is, or can be, warranted and some warrant for it would survive arbitrarily close scrutiny of its pedigree and arbitrarily extensive increments to, or other forms of improvement of, our information.

This admittedly vague characterization makes purely formal use of the notions of “state of information,” “improvement,” and so on. It's natural to wonder how more concrete yet generally applicable accounts of these notions might be given. But I do not think we need to take these issues on. It is enough for our purposes if the notion of superassertibility is *relatively* clear; clear, that is, relative to whatever notion of warranted assertion is in play in the particular discourse with which we may happen to be concerned. Provided, as in all cases that interest us there will be, there

are generally acknowledged standards of proper and improper assertion within the discourse, there must be sense to be attached to the idea of a statement that under certain circumstances meets the standards of proper assertion and then will or would continue to do so unless the considerations that led to its downfall were open to objection in some way. In short, wherever our discourse displays some measure of convergence about what is warrantably assertible, a corresponding notion of *superassertibility* has to be intelligible. This notion may be unclear in various respects, but they will be respects in which the relevant notion of warranted assertibility was already unclear.

So does superassertibility qualify as a potential truth property—does it satisfy the minimal platitudes? The issues here are actually quite subtle. Let's explore some of the twists. We already noted that superassertibility is, plausibly, both absolute and stable.<sup>32</sup> It is uncontroversial that it is potentially divergent in extension from assertibility proper. But it merits consideration whether superassertibility and assertibility coincide in normative force. And the question, anyway, is not merely whether superassertibility has these features but whether they issue in the right kind of way from its sustaining the key platitudes.

Let's focus on the Equivalence Schema. Can a supporter of superassertibility argue compellingly for the validity of (E<sup>s</sup>):

(E<sup>s</sup>) It is superassertible that *P* if and only if *P*.

If he can, then, as briefly noted above, that will arguably settle the matters of Transparency, Correspondence, and Contrast. The commutativity of superassertibility and negation—the analogue of (NE) for superassertibility—will likewise be a consequence.<sup>33</sup>

The matter may seem easily resolved, at least to anyone sympathetic to the idea that for a wide class of admissible substitutions for "*P*," it may be that *P* although no evidence is available to that effect. Such a theorist will want to object that (E<sup>s</sup>) cannot be valid, since it conflates right across the board the obtaining of a certain kind of high-grade evidence for *P* with the obtaining of the fact. A suitably chosen proposition—Goldbach's conjecture, say—may be undetectably true, and hence not superassertible, and a suitably chosen superassertible proposition—perhaps that we are not brains-in-a-vat—may be undetectably false. Since (E<sup>s</sup>) is hostage to

counterexample, so not a priori true, superassertibility has no case to be a truth property.

But the supporter of superassertibility may rejoin that, quite apart from any doubt about the realism on which it depends, there is something unsatisfactory about the *shape* of this objection. Its claim is that there is no assurance that there are no counterexamples to (E<sup>s</sup>). But what does it take a counterexample to be? Is it a true proposition that may not truly be claimed to be superassertible? In that case the objection asserts, in effect, that superassertibility potentially lacks, but as a putative truth property ought to be guaranteed to have, the property of generating a valid equivalence when substituted for “?” in the schema (F):

(F) It is true that it is ? that *P* iff. it is true that *P*

However, (F) contains two mentions of a truth property, which, if interpreted as presupposed by the objection, has to be understood as *distinct* from superassertibility. If that doesn’t seem evident, reflect that while—to one in the cast of mind that fuels the objection—it is a possibility that Goldbach’s conjecture be true without it being true that it is superassertible (provable), it certainly isn’t evident that the conjecture might be *superassertible* without it being superassertible that it is. But if there really can be, as minimalism suggests, a *plurality* of truth properties, qualifying as such by satisfying certain general principles, it is only to be expected that an *illusion* of failure may be created by selective interpretations of “true” as it occurs within those principles. It is as if someone were to argue that physical necessity fails to qualify as a genuine notion of necessity on the grounds that it fails to satisfy the principle  $\text{Necessarily } (A \leftrightarrow B) \models \text{Necessarily } (A) \leftrightarrow \text{Necessarily } (B)$

and were then to try to back up that contention by selectively interpreting the final occurrence of “Necessarily” in terms of *logical* necessity. If we wish to determine whether there are counterexamples to (E<sup>s</sup>), the proper question to put, the friend of superassertibility contends, is not whether superassertibility satisfies (F), but rather whether it satisfies what results when the two tendentious occurrences of “true” are replaced by ones of “?”:

(G) It is ? that it is ? that *P* if and only if it is ? that *P*



The question is, in effect, whether, whenever it is superassertible that  $P$ , it is superassertible that it is so, and vice versa.

Can we arbitrate this exchange? What is suspect about the shape of the original objection can be put like this. If any genuine truth property has to validate (i.e., satisfy a priori) the Equivalence Schema, then clearly, *distinct* truth properties can operate over a single discourse (or range of propositions) only if they are a priori coextensive. Plainly, then, no predicate  $F$  can express such a property in a discourse in which it is made to function alongside another predicate  $G$  that is already assumed both to validate the Equivalence Schema and to be potentially divergent in extension from  $F$ .<sup>34</sup> The original objection is therefore cogent only to this extent: to show that a discourse is governed by an evidentially unconstrained notion of truth is, for that reason, to show that superassertibility is not a truth property *for that discourse*.<sup>35</sup> But no *global* conclusion is licensed. We have to distinguish the questions (i) whether a predicate's content would enable it, under certain conditions, to function as expressing a truth property; (ii) whether, if so, the relevant conditions are met by any particular discourse; and (iii) whether they are met globally. The objection, drawing as it does on a range of examples where it is thought especially plausible that truth is evidentially unconstrained, is properly targeted against the claim of superassertibility on a positive answer to (iii). But in failing to make any distinction among the three questions, it implicitly begs the other two.

There is, however, on the other side, a similar oversimplification in the suggestion that "*the* proper question to put" is, in effect, whether

(G<sup>s</sup>) It is superassertible that it is superassertible that  $P$  iff. it is superassertible that  $P$

holds a priori. The right perspective, rather, is this. In the presence of the Equivalence Schema, counterexamples to (E<sup>s</sup>) are indeed all and only cases where

(F<sup>s</sup>) It is true that it is superassertible that  $P$  iff. it is true that  $P$

also breaks down. So if (G<sup>s</sup>) is valid, then we know that there can be no such counterexamples, and hence that (E<sup>s</sup>) is valid, provided, but only provided *no competitor truth-property* operates alongside super-

assertibility—no predicate, that is, that validates the Equivalence Schema but whose coextensiveness with superassertibility is not guaranteed a priori. If there is a competitor in operation, ( $F^s$ ) may fail when its occurrences of “true” are suitably interpreted, even if ( $G^s$ ) is valid without restriction on “P.” If there is no competitor, ( $G^s$ ) and ( $F^s$ ) stand or fall together. The status of ( $G^s$ ) is thus highly germane to question (i). If counterexamples to it cannot be excluded a priori, then there will be no general assurance that superassertibility can function as a truth property even when we give it the fullest elbowroom, as it were—even when we make no initial assumption that a competitor is operating over the discourse. On the other hand, if counterexamples to ( $G^s$ ) can be excluded a priori irrespective of the range of “P,” then we can return a positive answer to question (i), and the answers to questions (ii) and (iii) will then depend on whether and how widely competitor truth properties should be regarded as in operation.

So is ( $G^s$ ) unrestrictedly valid? We may return a positive answer if it can be shown that to have warrant for  $P$  is to have warrant for the claim that  $P$  is superassertible, and conversely.<sup>36</sup> The latter direction seems unproblematic. If we have reason to regard a statement as superassertible, then we have reason to think that some (in principle accessible) state of information will stably justify the statement, no matter how added to or otherwise improved. And having reason to think that such a state of information exists is plausibly taken to have the same probative force as actually being in the state of information in question. For instance, proving that a (canonical) proof of a particular statement can be constructed is, as far as probative force is concerned, as good as constructing the proof; and there seems no reason why the point should not survive generalization to the general run of cases where we are concerned with defeasible grounds rather than conclusive ones like mathematical proof.

What is less clear is that to have warrant to assert a statement must be to have warrant to regard it as superassertible. Doubtless, warrant to assert  $P$  cannot coexist with warrant to *deny* that  $P$  is superassertible, since that would be to have warrant to think that the present case for  $P$  would be defeated if we pressed matters sufficiently far, and again, that seems as much as to defeat it already. But the question to ask is, rather, whether warrant to assert  $P$  can coexist with lack of warrant to regard it

as superassertible—whether one can coherently combine *agnosticism* about *P*'s superassertibility with regarding a present case for asserting it as sufficient.

I'll outline an argument that the mooted combination of attitudes is *not* coherent, that it is precluded by certain quite basic elements in our ordinary conception of what justification for a statement or, equivalently, warrant for a belief involves. The elements involved are three. The first is that epistemic warranty does not have a sell-by date—what I am warranted in believing I remain warranted in believing *sine die* unless I acquire defeating collateral information. The second is that in warrantably believing any statement *P*, a subject is thereby warranted in believing that a sound investigation, to whatever extent one is possible, would bear her out. The third I shall introduce in a moment.

Suppose I warrantably believe that *P*. Now, what counts as warrant to believe a particular statement varies, of course, as a function of time, place, and background information. So what counts as corroboration of *P* for me if I return my attention to the matter in a year's time, say, may comprise very different considerations to those which warrant my present belief. However, by the first of the two assumptions, I will then be warranted, *ceteris paribus*, in believing *P*; and by the second, I will thereby be entitled to expect whatever sound considerations are then available to me to be corroborative just in virtue of the warrant I possessed a year before.

That establishes a conditional: if I am warranted in believing *P* now, then, if I acquire no other relevant information in the meantime, I will be warranted in future in expecting then-available, sound considerations to bear *P* out. But this conditional is something that I may take myself to know now. So whenever I know its antecedent—which, as remarked earlier, I can whenever it is true, since possession of warrant should be a decidable matter—I can know that in any case where I acquire no further relevant information in the interim, certain expectations will be warranted in future. But to know that certain beliefs will be warranted in the future is, only provided one has no present reason to view them as wrong, to be warranted in holding them now. This is the third element in our ordinary conception of justification advertised above: the firm *promise* of justification for what one has no reason to doubt is already

justification. So to be warranted in believing  $P$  involves having justification for believing that any subsequent, soundly conducted investigation, prior to which one has acquired no further relevant additional information, will corroborate  $P$ .

This is *close* to the desired result but doesn't quite get it. What would suffice to justify the claim that  $P$  is superassertible is warrant for the claim that *any* improvement,  $I^*$ , of my present state of information,  $I$ , will justify  $P$ . But what the foregoing establishes is only that if I am warranted in believing  $P$ , then I am warranted in claiming that any such  $I^*$  *prior to which I have acquired no further relevant additional information* will justify  $P$ . So there is a gap. But perhaps we can eliminate it given the third assumption mooted at the end of the preceding paragraph. Say that a later state of information  $I^*$  is *first-time  $P$ -incremental* on an earlier one  $I$  for a given thinker just if prior to possessing  $I^*$ , she has no  $P$ -relevant information that she did not possess in  $I$ . So our result above was that if I am warranted in believing  $P$  in  $I$ , then I am warranted in thinking that each  $I^*$  that is first-time  $P$ -incremental on  $I$  will likewise warrant  $P$ . And now, in order to extend this result to arbitrary improvements  $I^*$  of my present state of information  $I^*$ , it suffices to reflect that if  $I^*$  is not first-time  $P$ -incremental on  $I$ , then it must be the terminus of a finite chain,  $\langle I, I^2, \dots, I^* \rangle$ , each element of which is first-time  $P$ -incremental on its immediate predecessor. (The point is simply that no matter what  $P$ -relevant information I gather between  $I$  and  $I^*$ , there has to be a first state of information in which I possess each particular item in it.) Reflect then that, by the result of the previous paragraph, in each  $I^k$  in which I am warranted in believing  $P$ , I will be warranted in believing that I will be warranted in believing  $P$  in  $I^{k+1}$ . I can know this in  $I$  and hence infer that I am warranted in believing that in  $I^2$ , I will be warranted in believing that in  $I^3$ , I will be warranted in believing ... that I will be warranted in believing  $P$  in  $I^*$ . Application of the third assumption will then let me simplify to "If I am warranted in believing  $P$  in  $I$ , then I will be warranted in believing  $P$  in any improved state of information  $I^*$ ."

The contention that  $(G^s)$  holds a priori, without restriction on the range of " $P$ ," is thus very much in play, but I leave it to the reader to satisfy herself of the premisses and detail of this argument, which will

bear a more rigorous examination.<sup>37</sup> In general, though, it is hard to see how the making of warranted assertions, and the avoidance of unwarranted ones, could have any distinctive point or consequence unless warrant is taken per se to license expectations about the favourable character of subsequent states of information.<sup>38</sup>

## 5 Superassertibility as a Model of Truth

We now need to observe, finally, that it is actually not necessary, in order for superassertibility to qualify as a truth property, that it validate the platitudes unconditionally. It will be of no less significance if superassertibility turns out to validate the basic platitudes only subject to certain *additional assumptions* that, consistently with the platitudes, hold a priori for a particular discourse. Such a finding would put us in a position to say that, whether or not the platitudes are analytic of superassertibility when all occurrences of “true” are so interpreted, it is at least a possible *model* of them: it can be shown to have the features they collectively articulate when they are augmented with suppositions on whose status the platitudes themselves are silent.

How does the inquiry fare if we let it take this direction? One way of pursuing the matter begins by asking what is the relation between superassertibility and knowledge. It would be a tall order to argue unrestrictedly that whatever is superassertible can be known, not merely because one would have to vanquish the metaphysical-realist notion that even an empirically unimprovable theory might simply be mistaken, but perhaps more seriously, because the superassertibility of a statement carries no implication about the *strength* of the available evidence, which, though positive, may be enduringly weak. By contrast, it seems to me a highly intuitive claim that anything we can know is superassertible. Admittedly, this will not be so on any reliabilist conception of knowledge sufficiently extreme to abrogate all connection between knowledge and the possession of reason to believe. On such a view, one can know that *P* just by being a dispositionally reliable litmus of whether or not *P*, even if one has nothing whatever to say in support or explanation of one’s believing or disbelieving *P*. But on any view according to which knowledge requires at least some backup with reasons, that is, with asserti-

bility, it is surely going to require superassertibility too. I do not deny that in suitable circumstances an agent may know something on the basis of information that can in fact be defeated. But if his knowledge claim is not to be undermined by the availability of such defeating information, it is surely required that the negative effect of that information, once acquired, could itself be stably overturned.

Doubtless, the matter needs more discussion. But let me propose (K) as analytic of the concepts of knowledge and superassertibility:

(K)  $P$  is knowable  $\rightarrow P$  is superassertible

And now suppose we are dealing with a discourse in which, as we conceive, it is guaranteed a priori that each statable truth can, in favorable circumstances, be recognized as such—a discourse for which we can make nothing of the idea that truth might lie beyond all possibility of acknowledgement. Comic and, on a wide class of views about it, moral discourse are each, for instance, in this situation: there seems no sense to be attached to the idea that the comedy of a situation might elude the appreciation even of the most fortunately situated judge, or that the moral significance of an act might lie beyond human recognition, even in principle.<sup>39</sup> In any case, suppose that, for each assertoric content,  $P$ , in some germane class, we have it a priori that:

(L)  $P \leftrightarrow P$  is knowable

Had we the converse of (K),

$P$  is superassertible  $\rightarrow P$  is knowable,

the validity for the discourse concerned of the Equivalence Schema for superassertibility,

(E<sup>s</sup>) It is superassertible that  $P$  if and only if  $P$ ,

would, of course, be immediate. But we can skin the cat without appeal to the converse of (K) provided we are entitled to assume one half, as it were, of the commutativity of superassertibility and negation, specifically the direction from

It is superassertible that [not  $P$ ]

to

Not [ $P$  is superassertible].<sup>40</sup>

This principle is equivalent to the inconsistency of the supposition that  $P$  and its negation might both be superassertible, and is therefore uncontroversial so long any two states of information are conceived as mutually accessible and warrant is so conceived that no state of information can warrant contradictory claims.

With this lemma in place, it is easy to see that  $(E^s)$  is good. What needs to be shown is that

$P$ , and  $P$  is not superassertible.

and

$P$  is superassertible, and Not- $P$ .

are contradictory, just as are “ $P$  and  $P$  is not true”, and “ $P$  is true and not- $P$ ”. For the first, merely reflect that if  $P$  then, by (L),  $P$  is knowable; and if  $P$  is knowable, then, by (K),  $P$  is superassertible. For the second, reflect that, by the same moves, if Not- $P$ , then Not- $P$  is superassertible, and hence by the commutativity lemma, that it’s not the case that  $P$  is superassertible, contradicting the first conjunct. Thus, granted the a priori link between knowability and superassertibility postulated by (K), it follows, for any set of contents that sustain (L) a priori, that the assertion of any of these contents is a commitment to its superassertibility and the assertion of its superassertibility is a commitment to (rejecting any denial of) the content.

Plausibly, then, for discourses all of whose contents are in that case, superassertibility satisfies the Equivalence Schema and, in the light of earlier considerations, thus plausibly presents a model of the basic platitudes.<sup>41</sup> And if what I said about the essential appreciability of the moral and the comic is correct, a presumption is established that moral and comic truth can be taken as species of superassertibility.<sup>42</sup>

One interesting effect is the perspective in which the semantical anti-realism is now placed that generalizes Michael Dummett’s interpretation of mathematical intuitionism. Dummett’s antirealist, inspired by considerations concerning the acquisition and manifestation of understanding, contends that if the meaning of a statement is to be regarded as determined by its truth conditions, then truth cannot outrun our ability (in principle) to know. But then the thesis is that assumption (L), the equiv-

alence of “*P*” and “*P* is knowable,” holds globally for *all* intelligible assertoric contents. So, granted (K), the semantical antirealist contention becomes, in effect, that truth behaves, or ought to behave *everywhere* in a fashion that allows it to be construed as superassertibility. And to respond to the manifestation and acquisition arguments will be to explain how the currency of a notion of truth that cannot be modeled in terms of superassertibility is distinctively displayed in certain aspects of our linguistic practice, and how such a conception of truth might be arrived at in the first place.

This seems to me a helpful perspective on the Dummettian debate. Semantical antirealism now distances itself from the almost certainly doomed project of attempting a meaning-theory that proceeds in terms of an indexical notion of assertibility. Instead, it avails itself of a notion of truth, contrasting with assertibility, and an associated truth-conditional conception of meaning. But it can do this only because superassertibility is, as any antirealistically acceptable notion of truth must be, an essentially epistemically constrained notion—for if *P* is superassertible, it must be possible to alight on the (de facto) indefeasible state of information that makes it so, and then to accumulate inductive grounds for identifying it as such.

For the purposes of pragmatism, for its part, the crucial reflection is that superassertibility is, in a clear sense, an *internal* property of the statements of a discourse—a projection, merely, of the standards, whatever they are, that actually inform belief formation and assertion within the discourse. It supplies no external norm—in a way that truth is classically supposed to do—against which our ordinary standards might themselves be measured *sub specie Dei* and might rate as adequate or inadequate. Rather, the way in which it is fashioned from our actual practices of assessment renders superassertibility as well equipped to express the aspiration for a developed pragmatist conception of truth as any other candidate known to me. If it seems to distort our thinking about truth in particular regions of discourse to conceive it in such terms, that, it seems to me, will be a measure of the local unnaturalness of pragmatism itself.



## Notes

This paper, at Michael Lynch's suggestion, revisits some of the arguments and themes of chapters 1 and 2 of my *Truth and Objectivity* (Cambridge: Harvard University Press, 1993.) I am grateful to Michael Lynch for giving me the opportunity to present these ideas to the readership of the present volume and for helpful suggestions about what best to include. Sven Rosenkranz also gave me extremely detailed and helpful suggestions about both contents and structure. My thanks to Harvard University Press for permission to include excerpts from *Truth and Objectivity* and to the University of Calgary Press for permission to include passages from my more recent paper "Truth: A Traditional Debate Reviewed" (*Canadian Journal of Philosophy*, suppl. vol. 24 [1998]: 31–74; reprinted in *Truth*, ed. Simon Blackburn and Keith Simmons [Oxford University Press, 1999]; the official dates notwithstanding, first published in German in Matthias Vogel and Lutz Wingert, eds., *Unsere Welt gegeben oder gemacht? Menschliches Erkennen zwischen Entdeckung und Konstruktion* [Frankfurt am Main: Suhrkamp, 1999]). The present paper was completed during my tenure of a Leverhulme Personal Research Professorship. I gratefully acknowledge the support of the Leverhulme Trust.

1. That is, the Equivalence Schema and the Disquotational Schema yield instances whose truth is knowable a priori by anyone who is in a position to understand them. As is familiar, the right-to-left directions of these equivalences become contestable if truth-value gaps or many truth values are admitted. This complication is pursued in discussion note 1 of chapter 2 of *Truth and Objectivity*. But I do not think that any deflationist should go out of her way to accommodate it, since rejection of the right-to-left direction of the Equivalence Schema flies in the face of what would seem to be an absolutely basic and constitutive characteristic of the notion of truth, that  $P$  and "It is true that  $P$ " are, as it were, attitudinally equivalent: that any attitude to the proposition that  $P$ —belief, hope, doubt, desire, fear, etc.—is tantamount to the same attitude to its truth. For if that's accepted, and if it is granted that any reservation about a conditional has to involve the taking of *differential* attitudes to its antecedent and consequent, then there simply can be no coherent reservation about  $P \rightarrow$  it is true that  $P$ .

2. It's an unhappy situation that the leading contemporary theorist of deflationism, Paul Horwich, uses both "minimalism" and "deflationism" to characterize his view. However, both his use of "minimalism" and my contrasting one are now entrenched. Probably nobody is confused.

3. This view is contrary to what is suggested by Horwich, *Truth*, 2nd edition (Oxford: Clarendon Press, 1998), pp. 143–144. Horwich there seems to conflate the substantiality of a property with the feasibility of what he calls a "theory of constitution" for this property, i.e., a theory that identifies this property by means of a noncircular equation of the form " $x$  is true iff  $x$  is  $F$ ," where " $F$ " is replaced by a predicate that does not contain any semantic terms, a fortiori no cognates of "is true." But that just seems to be a prejudice. It is evident from the example of scientific-theoretical predicates, for instance, that there can be no compelling reason to tie expression of a substantial property to explicit definability.

4. There are deflationists who go so far as to deny that “is true” is a genuine predicate at all, but most deflationists are ready to concede that there is such a thing as the concept of truth. A deflationist proposal of the first kind can be found in Grover et al., “A Prosentential Theory of Truth,” *Philosophical Studies* 27 (1975): 73–125.

5. That is, the property of having “true” correctly predicable of them. This is presumably what Horwich has in mind when he says that truth denotes a property in the sense in which “every term that functions logically as a predicate stands for a property” (*Truth*, 2nd edition, pp. 141–142).

6. See Hilary Putnam, “On Properties,” in his *Mathematics, Matter, and Method* (Cambridge: Cambridge University Press, 1975), pp. 305–322.

7. Thus, the minimalist opposes Horwich’s suggestion that truth presents a special case in that an account of the property (or properties) denoted just coincides with an account of the concept that does the denoting. See his *Truth*, 2nd edition, p. 136.

8. On Horwich’s interpretation of “substantive property,” such reducibility is precisely a necessary condition for a property to be substantive. His suggestion that minimalism (in my sense) is based on the idea that truth is substantive on this understanding thus misconceives the position. See his *Truth*, 2nd edition, pp. 142–143.

9. Proof: derive the two biconditionals one gets from (ES) by respectively negating both its halves and taking “not-*P*” for “*P*.” Transitivity of the biconditional then yields (NE).

10. If they were necessarily coextensive, the Negation Equivalence would have to hold for both if for either.

To offset misunderstanding, two points merit emphasis. First, warranted assertibility is here understood to be a notion that is always relativized to a particular state of information. If no such state of information is explicitly mentioned, claims involving this notion will always be understood to relate to the present state of information. Second, the modality involved in “warranted assertibility” does not signify the potential possession of warrants for an assertion, but the actual possession of warrants for a potential assertion. So in particular, merely provable mathematical statements, for which we so far have no proof, do not qualify as warrantably assertible. I believe a confusion of this distinction drives the criticisms in Neil Tennant’s paper “On Negation, Truth, and Warranted Assertibility,” *Analysis* 55 (1995): 98–104.

11. Thus, for instance Horwich, *Truth*, 2nd edition, pp. 20–23, 139–140.

12. This is, of course, by no means the end of the dialectic. A supporter of the project of Robert Brandom’s compendious *Making It Explicit* (Cambridge: Harvard University Press, 1994) will believe that a suitable account of assertoric content—one sustaining the contrast between the proposition that *P* and the proposition that that proposition is assertible—can be constructed out of truth-free materials, as it were. And in his recent book *Meaning* (Oxford: Clarendon

Press, 1998), Horwich himself tries—as he must—to develop a general account of meaning in which truth plays no explanatory part. I cannot pursue the problems with these approaches here. My own view is that the best deflationist response to the inflationary argument is to concede its immediate conclusion but insist that it shows no more than that the concept of truth is indeed of a dimension of (substantial) success and failure, distinct from warrant, for *each particular* proposition, but that there still need be no *single* thing in which, for any two propositions, such success or failure consists. This is indeed one way of taking the “stock-in-trade” response reviewed above. For pursuit of the issue at least some distance beyond this point, see my “Truth: A Traditional Debate Reviewed,” section IV.

13. Which of these forms of opacity goes with the very concept of truth is, of course, contentious, but not that some do.

14. For elaboration of this claim, see my *Truth and Objectivity*, pp. 24–27.

15. Readers familiar with Michael Smith’s work will note a point of contact here with the conception of a *network analysis*, which he derives from Ramsey and Lewis (see in particular chapter 2, section 10, of Smith’s *The Moral problem*, Oxford: Basil Blackwell, 1994). The principal contrast with the approach to truth here canvassed is that a network analysis has to be based on a comprehensive set of platitudes whose conjunction so constrains the target concept that the replacement within those platitudes of all expressions for that concept by a variable and its binding by the description operator results in a definite description that is at the service of an analytically true identity:

*F*-ness is the property  $\Phi$  such that  $\{\dots\Phi\dots\&\dots\Phi\dots\&\dots\}$

This effectively supplies a reductive analysis of the concept *F*. An analytical theory, by contrast, need not—though it may—subserve the construction of such an analytically true identity.

16. C. S. Peirce, *Collected Papers*, ed. C. Hartshorne and P. Weiss (Cambridge: Harvard University Press, 1935), vol. 8, p. 139. [See chap. 8.—Ed.]

17. We may take it that this is the notion that is now standardly called “assertibility.” Putnam’s grounds for the rejection are two: first, that truth is, plausibly, timeless, whereas warranted assertibility varies as a function of the state of information (“Truth is supposed to be a property of a statement that cannot be lost, whereas justification can be lost”), and second, that assertoric warrant is, whereas truth is not, a matter of *degree*. Recall that I incorporated these points into the platitudes listed above.

18. *Reason, Truth, and History*, p. 56. [See chap. 11.—Ed.]

19. For a review of the contrast between Putnam’s “middle” and most recent views, see my “Truth as Not Wholly Non-epistemic: Putnam’s Peregrinations,” *Journal of Philosophy* 97 (2000): 335–364.

20. Hilary Putnam, *Realism with a Human Face*, ed. James Conant (Cambridge: Harvard University Press, 1990).

21. *Realism with a Human Face*, Preface, p. viii.

22. *Realism with a Human Face*, Preface, P. viii.

23. In making this distinction between the Peircean and sometime Putnamian conceptions, I intend no judgement about whether it is finally stable. As noted, Putnam's intention was that truth, as he informally elucidates it, is, in contrast to warrant, to be a stable property of propositions across time and a property that is absolute, that is not applicable in varying degree. Plainly, this intention can be fulfilled only if to have warrant for a proposition under "epistemically ideal conditions" (however that phrase be interpreted) involves having a case for it that *cannot be defeated* (else we wouldn't have stability) *or improved* (else we wouldn't have absoluteness) *by any further information*. And the only way of ensuring that both points are met would seem to be to require that circumstances count as epistemically ideal (or topic-specifically sufficiently good) with respect to a particular statement just in case *no further information relevant to a verdict on it exists to be had*.

The force of that idea obviously depends on what "relevant" should mean in such a context. In fact, though, it is difficult to see that the term can impose any real restriction at all. For, as is very familiar, warrant is a highly systematic, holistic property of beliefs: the status of a body of information as support for a particular belief turns not simply on the character of the information and the content of the belief but on what beliefs are held as background. A flash of grey glimpsed in the woods may be evidence of the presence of a squirrel if you take yourself to be in New Jersey, say, but of a wood pigeon if you take yourself to be in Scotland. It is no exaggeration to say that any piece of information may, in the context of an appropriate epistemic background, be relevant to any particular belief. How, in consequence, are we to understand the idea of possessing *all* information relevant to a particular proposition? Doesn't it just have to mean possessing *all empirical information*, period? In this way, and notwithstanding his protestations to the contrary, Putnam's intentionally less extreme proposal may seem to slide inevitably toward the Peircean. But I make no assumption about this in what follows.

24. This requirement is superfluous, presumably, since a statement does not count as justified, in any sense that concerns us, unless the case in its favor dominates anything that counts in favor of its negation.

25. For further discussion, see *Truth and Objectivity*, chapter 2, discussion note 1.

26. Alvin Plantinga, "How to Be an Anti-realist," *Proceedings and Addresses of the American Philosophical Association* 56 (1982): 47–70. Plantinga believed he had Putnam in his sights as well, but there are some issues about that, as we will see (though he would be right in any case if the suspicion expressed in note 23 is sound.)

27. Assuming—surely correctly—that a subjunctive conditional, no less than an indicative, is controverted by the actual truth of its antecedent and falsity of its consequent.

28. Plantinga's version of this argument exploits the S4 principle—that what is necessary is necessarily necessary—to derive the conclusion that the idealization  $Q$  holds of necessity. But the derivability of  $Q$ , unnecessitated, is quite bad enough. A proponent of the “Peircean” conception, or a coherence account of truth, certainly would not intend that the actual obtaining of epistemically ideal conditions, or the actual existence of a maximally coherent belief set, should be consequences of the account. Indeed, these conditions are precisely thought *not* to obtain—hence the counterfactual analysis.

29. A useful explicit discussion is Robert K. Shope's “The Conditional Fallacy in Contemporary Philosophy,” *Journal of Philosophy* 75 (1978): 397–413. The Conditional Fallacy is, of course, a crucial difficulty for certain classical forms of philosophical reductionism—behaviorism and phenomenism, for instance—but like another absolutely basic structural problem for such views, the holistic interdependencies discussed in Christopher Peacocke's *Holistic Explanation* (Oxford: Clarendon Press, 1978), seems never to have been clearly appreciated during the heyday of debate about them.

30. Both pragmatist conceptions also confront a distinct worry concerning the implicit assumption that epistemically ideal or topic-specifically sufficiently good circumstances are *unique*. Only if so can the proposed conceptions of truth ensure convergence of opinion under such circumstances. But given that the relation *is evidence for* is holistically conditioned by background empirical theory, what a priori obstacle is there to the possibility that *conflicting* sets of beliefs be arrived at under epistemically ideal or topic-specifically sufficiently good conditions as a result of theorists having successfully maintained distinct theoretical backgrounds throughout the information-gathering process—so that an opinion formed about a particular statement can vary as a function of the direction in which, so to say, the idealized circumstances are approached? This thought is amplified, in rather a different context, in chapter 4 of *Truth and Objectivity*.

31. See my *Realism, Meaning, and Truth* (Oxford: Basil Blackwell 1987), pp. 295–302, and *Truth and Objectivity*, *passim*.

32. In fact, it is stable provided the range of the “states of information” quantifier in its definition is stable. That's an assumption that would be questioned by, for instance, an antirealist about the past, or future, who contested whether we should think of the totality of states of affairs as eternal. But, of course, such an antirealist would regard the truth predicate as unstable in any case, so that, in the view of such a theorist, instability stemming from that source would not disqualify superassertibility as a truth predicate. For further reflections on the matter, see *Realism, Meaning, and Truth*, pp. 300–302.

33. Of the remaining platitudes, Embedding is presumably uncontroversial if all assertoric contents sustain it and all are apt to be superassertible. A degree of Opacity is likewise uncontroversial for superassertibility (though what degree of Opacity any truth property has to display is in any case likely to be a vexed question).

34. This is because it cannot be a priori that ( $P$  iff  $P$  is  $F$ ) if it is a priori that ( $P$  iff  $P$  is  $G$ ) but not a priori that ( $P$  is  $G$  iff  $P$  is  $F$ ).

35. Since, trivially, if  $P$  is superassertible, there has to be evidence for  $P$ .

36. For suppose that to have warrant for  $A$  is to have warrant for  $B$  and vice versa, but for *reductio*, that  $A$  is superassertible, while  $B$  is not. Let  $I$  be a total state of information in virtue of which  $A$  is superassertible, i.e.,  $I$  warrants  $A$  and so does any improvement  $I^*$  of  $I$ . By hypothesis,  $I$  also warrants  $B$ . Since  $B$  is not superassertible, there must therefore be some improvement  $I^*$  of  $I$  that fails to warrant  $B$ . Since any such  $I^*$  warrants  $A$ , the supposition is contradicted. This shows that coincidence in assertibility conditions suffices for a pair of statements both being superassertible if either is. So if " $P$ " and " $P$  is superassertible" have the same assertibility conditions, ( $G^s$ ) follows.

37. A beginning is made in discussion note 3 at the end of chapter 2 of *Truth and Objectivity*.

38. For further discussion of this general thought, see chapter 9, note 13, of *Realism, Meaning, and Truth* and the other passages in that book there referred to.

39. I prescind from the complication that the bearers of comic and moral predicates may be spatially or temporally remote. Naturally, modifiers of time and place throw up the same prima facie barriers to the acknowledgeability of comic, or moral truth, broadly conceived, as they pose for discourses in general. A similar point applies, of course, to quantification.

40. The other direction may easily be established by appeal to (K) and (L) as follows:

- (1) Not [ $P$  is superassertible] hypothesis
- (2) Not [ $P$  is knowable] 1 (by K)
- (3) Not  $P$  2 (by L)
- (4) It is knowable that [not  $P$ ] 3 (by L, 'not  $P$ /' $P$ )
- (5) It is superassertible that [not  $P$ ] 4 (by K)

A different argument for ( $E^s$ ) is presented in the Appendix to "Truth: A Traditional Debate Reviewed."

41. Such a conclusion could be drawn locally, of course, even if the general validity of (K) is rejected, provided that knowledge entails superassertibility in at least some discourses of which (L) is a priori true.

42. Only a presumption, though. A discourse that meets the conditions described, and so permits superassertibility to model the platitudes characteristic of truth, may yet have other features that impose differences between the two concepts. Getting clear about what such features could be is exactly what is involved in getting clear how realist/antirealist debate is possible after minimalism about truth is accepted on both sides.



---

## Contributors

**Linda Martín Alcoff** is Merideth Professor and Associate Professor of Philosophy at Syracuse University. She is most recently the author of *Real Knowing: New Versions of the Coherence Theory* and the editor of *Epistemology: The Big Questions*.

**W. P. Alston** is Emeritus Professor of Philosophy at Syracuse University. Professor Alston has made important contributions to epistemology, the philosophy of religion and the philosophy of language. His most recent books include *The Reliability of Sense Perception* and *A Realist Conception of Truth*.

**J. L. Austin**, the author of *Sense and Sensibilia* and *How to Do Things with Words* and was a leading figure in the “ordinary language” movement of philosophy prominent in the mid twentieth century.

**Brand Blanshard**, American absolute idealist, spent much of his distinguished career teaching at Yale University. His many books include *Reason and Analysis* and the multivolume *The Nature of Thought*.

**Marian David** is Associate Professor of Philosophy at the University of Notre Dame, working in metaphysics and the philosophy of mind. He is the author of *Correspondence and Disquotation*.

**Donald Davidson** is Willis S. and Marion Slusser Professor of Philosophy at the University of California at Berkeley. He is author of widely influential articles on language and mind, many of which are included in *Actions and Events* and *Inquiries into Truth and Interpretation*.

**Michael Devitt** is Professor of Philosophy at the City University of New York. He is the author of many important articles on language and realism and of *Realism and Truth* (2nd ed.) and *Coming to Our Senses*.

**Michael Dummett** is Emeritus Professor of Philosophy at the University of Oxford, where he was Wykeham Professor of Logic from 1972 to 1992. His influential work in the philosophy of language and mathematics includes *Frege: Philosophy of Language*, *The Logical Basis of Metaphysics*, *The Origins of Analytical Philosophy*, and *The Seas of Language*.

**Hartry Field**, Professor of Philosophy at New York University, has made numerous contributions to the philosophy of mathematics and the philosophy of



language. He is the author of *Science without Numbers: A Defense of Nominalism* and *Realism, Mathematics, and Modality*.

**Michel Foucault** was the leading philosopher of postmodernism. His many influential works include *The Archaeology of Knowledge*, *The Birth of the Clinic*, and the multivolume *History of Sexuality*.

**Dorothy Grover**, Professor Emeritus at the University of Illinois, Chicago, has contributed extensively to the philosophy of logic and language. She is the author of *The Prosentential Theory of Truth*.

**Anil Gupta** is Rudy Professor of Philosophy at Indiana University. He has been a leading figure for years in the field of philosophical logic. He is the author of *The Logic of Count Nouns* and (with N. Belnap) of *The Revision Theory of Truth*.

**Martin Heidegger** was the most influential and controversial German philosopher of the twentieth century. His books include *Being and Time*, *Discourse on Thinking*, and *The End of Philosophy*.

**Terence Horgan**, Professor of Philosophy at the University of Memphis, is the author (with J. Tienson) of *Connectionism and the Philosophy of Psychology* and of numerous articles on metaphysics and on the philosophy of mind. A collection of his papers, *Making Materialism: Mind and Its Place in Nature*, is forthcoming from Oxford University Press.

**Jennifer Hornsby** is Professor and Head of the School of Philosophy, Birkbeck College, University of London. Her publications include *Actions*, *Simple Mindfulness: A Defence of Naïve Naturalism in the Philosophy of Mind*, *Ethics: A Feminist Reader* (coedited with E. Frazer and S. Lovibond), and *The Cambridge Companion to Feminism in Philosophy* (coedited with M. Fricker).

**Paul Horwich** is Professor of Philosophy at the City University of New York. He is the author of *Probability and Evidence*, *Asymmetries in Time*, the much discussed defense of deflationism *Truth* (2nd ed.), and its companion *Meaning*.

**William James**, psychologist and philosopher, was one of the most influential American intellectuals of the late nineteenth and early twentieth centuries. His works include *The Principles of Psychology*, *The Will to Believe*, *Pragmatism: A New Name for Some Old Ways of Thinking*, and *The Meaning of Truth*.

**Michael P. Lynch** teaches philosophy at Connecticut College. He is most recently the author of *Truth in Context: An Essay on Pluralism and Objectivity*.

**Charles Peirce**, brilliant logician and founder of pragmatism, never held an academic post. His many papers are included in the eight-volume *Collected Papers of Charles Sanders Peirce*.

**Hilary Putnam**, Cogan University Professor at Harvard University, has made many important contributions to the philosophy of language, the philosophy of mind, and the field of mathematics. He is the author of *Reason, Truth, and History*, *Realism with a Human Face*, *Renewing Philosophy*, and *Pragmatism*.

**W. V. O. Quine** is Professor Emeritus at Harvard University. His many seminal contributions to the philosophy of language and logic include *From a Logical Point of View*, *Word and Object*, *Ontological Relativity and Other Essays*, and *The Pursuit of Truth*.

**F. P. Ramsey** was a Cambridge mathematician and philosopher who, before his untimely death at the age of twenty-seven, made important contributions to the philosophy of science and logic.

**Richard Rorty** is Professor of Comparative Literature at Stanford University and Professor of the Humanities Emeritus at the University of Virginia. His many books include *Philosophy and the Mirror of Nature*, *Truth and Progress*, and most recently, *Philosophy and Social Hope*.

**Bertrand Russell** won the Nobel Prize for Literature for his many contributions to philosophy and social criticism. One of the founders of analytic philosophy, his groundbreaking works includes *Principia Mathematica* (coauthored with A. N. Whitehead), *The Analysis of Mind*, and *Knowledge: Its Scope and Limits*.

**Scott Soames** is Professor of Philosophy at Princeton University. His books include *Understanding Truth*, *Propositions and Attitudes* (coedited with Nathan Salmon), and *Syntactic Argumentation and the Structure of English* (coauthored with David Perlmutter).

**Ernest Sosa** is the Romeo Elton Professor of Natural Theology and Professor of Philosophy at Brown University and Professor of Philosophy at Rutgers University. He is the author of many important articles in epistemology, and his books include *Knowledge in Perspective: Selected Essays in Epistemology* and *A Companion to Epistemology* (coedited with J. Dancy), and *A Companion to Metaphysics* (coedited with J. Kim).

**Peter Strawson**, Emeritus Professor of Philosophy at Oxford University, has made important contributions to the philosophy of language and metaphysics. His books include *The Bounds of Sense*, *Individuals: An Essay on Descriptive Metaphysics*, and *Analysis and Metaphysics*.

**Alfred Tarski**, mathematician, logician, and philosopher, spent the latter half of his life at the University of Berkeley, having emigrated from his native Poland. He is the author of numerous papers on logic and algebra, the more philosophical of which are included in *Logic, Semantics, and Metamathematics*.

**Ralph Walker**, a Fellow of Magdalen College, Oxford, is the editor of *Kant on Pure Reason* and *The Real in the Ideal*, and is the author of *Kant and The Coherence Theory of Truth*.

**Crispin Wright** is Bishop Wardlaw Professor of Philosophy at the University of St. Andrews and regular Visiting Professor at Columbia University. His books include *Wittgenstein on the Foundations of Mathematics*; *Realism, Meaning, and Truth* (2nd ed.); and the widely discussed *Truth and Objectivity*.



---

# Index

- Absolute idealism, 99, 103–105  
Accordance with reality, 296–299. *See also* Correspondence, nature of  
Adorno, T., 162  
Alcoff, Linda, 100–101, 159–182,  
292. *See also* Coherence theory of truth  
argues that coherence theory of truth  
is not antirealist, 166–172  
Alethic monism, 725–726  
Alethic realism, 9–14, 41–42,  
45–47. *See also* Alston, W. P.;  
Correspondence theory of truth;  
Realism  
Allen, Barry, 280 (n. 7), 291  
Alston, W. P., 11, 41–65, 145, 746,  
746 (n. 3). *See also* Alethic realism;  
Correspondence theory of truth;  
Realism  
argues alethic realism is consistent  
with metaphysical antirealism, 48  
objects to epistemic conceptions of  
truth, 57–64  
Antirealism (Dummett), 137–143,  
187, 246–249, 555 (n. 33), 556  
(n. 34), 618, 669, 705–707, 708,  
710–711, 712, 766, 780–781.  
*See also* Dummett, Michael;  
Metaphysical antirealism  
Aristotle, 9, 41, 108, 333, 357 (n. 6),  
439, 625, 626, 627, 628  
Armstrong, David, 280 (n. 4), 588,  
698, 703 (n. 10), 729, 746 (n. 8)  
Austin, J. L., 10, 25–40, 696. *See also*  
Correspondence theory of truth  
criticizes redundancy theory of truth,  
31–36  
criticizes Strawson, 36–37, 463  
Strawson's criticisms of, 447–469  
Ayer, A. J., 415 (n. 7), 484, 485, 488,  
490–491, 500 (nn. 9, 11), 502  
(n. 20), 606 (n. 11), 607 (n. 14),  
708  
Barnard, Robert, 746, 747 (n. 9)  
Beliefs as truth bearers, 18, 21–23, 26,  
125–126, 434, 652  
Belnap, N., 425, 486, 499 (n. 3), 552  
(n. 7), 556 (n. 39)  
Benacerraf, Paul, 394, 414  
Berger, Alan, 416 (n. 18)  
Berthelot, Terry, 746  
Bivalence, principle of, 135–137, 707,  
715, 765. *See also* Excluded middle,  
law of  
Blackburn, Simon, 134–135, 156  
(n. 11), 585, 588, 589, 606 (n. 11),  
607 (n. 14), 660 (n. 1), 667  
Blanshard, Brand, 54, 99–102, 103–  
121  
Bloomfield, Paul, 746  
Boghossian, Paul, 606 (nn. 9, 11), 607  
(n. 14)  
Bonjour, Laurence, 60, 101  
Bosanquet, B., 120–121 (n. 9), 444  
(n. 5)

- Bradley, F. H., 99, 114–115, 145, 157 (n. 18), 702 (n. 1)
- Brandom, Robert, 280 (n. 7), 286 (n. 60), 524 (n. 3), 583, 584, 604, 607 (n. 14), 609 (nn. 31, 34, 35), 677 (n. 6), 783 (n. 12)
- Buhler, Axel, 414
- Burgess, John, 414
- Camp, J., 425, 486, 499 (n. 3), 552 (n. 7)
- Candlish, S., 676 (n. 1), 676 (n. 5), 702 (n. 1)
- Carnap, R., 77, 360 (nn. 20, 21), 394 (nn. 2, 3), 415 (n. 7), 418 (n. 23), 552 (n. 1), 556 (n. 39), 608 (n. 25), 708, 719 (nn. 6, 9)
- Cartesianism, 376–377. *See also* Descartes, R.
- Cartwright, N., 170
- Cartwright, R., 689
- Casalegno, Paolo, 575 (n. 18)
- Causal theories of reference, 68, 253–255, 327, 384–385, 390–391, 404, 497, 603–604
- Chalmers, D., 557 (n. 49), 744, 747 (n. 13), 748 (n. 14)
- Chapuis, A., 557 (n. 49)
- Chastain, Charles, 524 (n. 14)
- Chisholm, Roderick, 685, 693, 702 (n. 1), 703 (n. 7)
- Clark, Peter, 281 (n. 18)
- Cognitive command (Wright), 271–273
- Cohen, L. J., 156 (n. 4)
- Coherence, 99, 107–109, 126–127 as test for truth, 103, 109–110
- Coherence theory of truth, 99–102, 103–122, 123–158, 159–182
- Alcuff on, 159–182
- Blanshard on, 103–122
- pure vs. impure, 146–147, 151–155
- and realism vs. antirealism, 132–143, 166–172
- and reality, 106–107, 109–110
- Russell's objections to, 19–20, 101, 125–126
- and thought, 104–105
- Walker on, 123–158
- Concept vs. property of truth, 3, 50, 733, 753–754
- Conditionals, 238–240
- Content, 683, 692–694. *See also* Propositions
- Convention T, 323, 335, 395 (n. 14), 627, 630, 706. *See also* Schema T; Tarski, A.
- Cook Wilson, J., 444 (n. 4)
- Correspondence, nature of, 9–12, 21–24, 28–31, 458–459, 471 (n. 9), 698–699, 700
- Correspondence theory of truth, 1–2, 9–14, 68–70, 77–79, 128, 145–146, 244–245, 251, 299–300, 333–334, 439–440, 485, 579, 617, 628, 642, 659, 667–668, 698–700, 724, 735, 746 (n. 2). *See also* Realism
- Alston on, 49–51, 145
- Austin's version of, 25–37, 447–469
- David on, 695–704
- Devitt on, 601–605
- Field on, 487–489, 494–495
- Frege's objection to, 665–666
- Horgan's theory of, 70–77
- objections to, 12–13, 51–57, 77–79, 149–151, 300–301, 451–452, 458–459, 467–468, 470 (n. 7), 475, 643–644, 646, 665–666, 724
- Russell's version of, 20–24
- Dalton, Eric, 556 (n. 46), 557 (n. 49)
- Dancy, Jonathan, 161
- Darwin, Charles, 277, 280 (n. 4)
- David, Marian, 429 (n. 2), 524 (n. 8), 557 (n. 49), 599–600, 605, 617, 678 (n. 10), 680 (n. 16), 683–704, 746
- Davidson, Donald, 52, 56, 190, 261–264, 273, 275–276, 279 (n. 10), 280 (n. 18), 281 (n. 19), 281–282

- (n. 23), 284 (nn. 45, 50, 51), 285 (nn. 52, 60), 369, 394, 417 (n. 23), 429 (n. 5), 540, 556 (n. 35), 561, 562, 643–644, 645, 673, 678 (n. 8), 678 (n. 11), 679 (nn. 12, 13), 720 (n. 16)
- criticizes Horwich, 632–636
- objects to deflationism, 427, 561–565, 574 (n. 8), 575 (n. 14), 630–636
- offers theory of empirical content of truth concept, 636–639
- and truth as indefinable, 615, 625–628, 636, 661 (n. 8) (*see also* Primitivism about truth)
- Davies, M., 677 (n. 6)
- Deflationism, 5–6, 260, 265–268, 352–353, 421–422, 508–509, 579–601, 629–636, 675, 706, 707, 708, 716, 751–759, 782 (nn. 2, 3), 783 (n. 4). *See also* Disquotationalism; Minimalism (Horwich); Performative theory of truth; Prosentential theory; Redundancy theory of truth
- Davidson's objections to, 630–636
- as denying that truth has a causal/explanatory role, 535–538, 590–591
- as denying that truth has a nature, 5–6, 421, 589–590
- as denying that truth is a (substantive) property, 422, 508, 572 (n. 1), 587–588, 646–647
- distinguished from correspondence theory of truth, 488–495, 580–601
- Dummett's objections to, 232–237, 571
- Field's definition of, 488–495
- Gupta's critique of, 527–556
- Gupta's definition of, 528
- Horwich's definition of, 572 (n. 1)
- Putnam's criticisms of, 707–710
- related to nonfactualism (Devitt), 585–592
- related to noncognitivism, 649
- Sosa's criticisms of, 646–653
- Wright's argument against, 754–759
- Delancy, C., 557 (n. 49)
- Dennett, Daniel, 280 (n. 4)
- Denotation, 367, 368, 370, 372, 382–383, 385–388. *See also* Reference primitive (Field), 367–368, 370
- Descartes, R., 116, 194–196, 280 (n. 4)
- Devitt, Michael, 394, 428–429, 508, 524 (n. 16), 579–611, 678 (n. 9), 724, 746 (n. 2)
- distinguishes deflationism from correspondence theory, 580–601
- presents correspondence theory of truth, 601–605
- sees deflationism as similar to nonfactualism, 585–592
- Dewey, J., 162, 211, 226, 259–260, 276, 277
- Diamond, Cora, 720 (n. 13)
- Disquotational schema. *See* Schema T
- Disquotationalism, 260, 424–425, 529–530, 529, 530, 572 (n. 1). *See also* Deflationism
- Field on, 425, 484–488
- Gupta's criticisms of, 529–552
- Quine on, 424, 473–481, 630–632
- Dodd, J., 676 (n. 5)
- Dretske, Fred, 279–280 (n. 4)
- Ducasse, C. J., 702 (n. 1)
- Dummett, Michael, 52, 59, 137–143, 187–188, 229–251, 269, 389, 417 (n. 23), 527, 540, 542, 579, 635, 639, 642, 678 (n. 9), 679 (n. 15), 719 (n. 8), 723, 780
- and antirealism, 137–143, 187, 246–249, 555 (n. 33), 556 (n. 34), 618, 669, 705–707, 708, 766, 780–781
- objects to deflationism, 232–237, 571
- objects to realism, 14 (n. 1), 246–248
- Dworkin, Ronald, 736, 738
- Eliminativism about truth, 579
- Epiphenomenalism, 645

- Epistemic theories of truth, 57–65, 71, 642, 724, 761–781. *See also*  
 Antirealism; Coherence theory of truth; Pragmatist theories of truth; Putnam, Hilary; Verificationism
- Equivalence schema, 559, 560, 569–570, 583, 592–594, 632, 650, 687, 728, 751, 755, 756, 757, 758, 765, 768, 772, 774, 780, 782 (n. 1). *See also* Schema T
- Etchemendy, John, 553 (n. 14), 557 (n. 49), 628
- Eternal sentences as truth bearers (Quine), 474
- Ethics, truth in, 85–90
- Evans, Gareth, 272
- Excluded middle, law of, 235–236, 248, 707–708, 766
- External perspective (Putnam), 231–252. *See also* Metaphysical realism; Realism
- External realism (Sosa), 643. *See also* Metaphysical realism; Realism
- Facts, 9–10, 51, 475, 643–644, 665, 684, 690–691, 696–698, 699–700, 731, 732  
 Austin on, 28–30  
 Russell on, 23–24  
 Strawson on, 450–458, 466–468, 470 (n. 6)
- Falsehood, 18, 20, 234–235, 307–312, 464, 617, 689–690, 699
- Farin, Ingo, 557 (n. 49)
- Field, Hartry, 10, 326–327, 365–396, 429 (n. 3), 508, 523–524 (n. 8), 532, 535, 554 (n. 20), 555 (n. 24), 572, 596, 605, 609 (n. 33), 725, 746, 746 (n. 2)  
 and deflationism, 425, 427, 483–504, 565, 576 (n. 19) (*see also* Disquotationalism)  
 sees theory of truth as theory of truth conditions, 483–484  
 Soames's criticisms of, 403–408, 409, 411–413  
 on Tarski, 326–327, 365–396, 495–499
- Fletcher, Charles, 746
- Fodor, J., 720 (n. 16)
- Foley, Richard, 60
- Foucault, Michel, 163, 291–292, 317–319
- Freedom, 303–307
- Frege, G., 149–150, 229–231, 233, 394 (n. 10), 421, 485, 561, 564, 625, 626, 658, 665–667, 677 (n. 6), 684, 692, 702 (n. 1), 703 (n. 6), 714
- Functionalism, 727–729, 738–739
- Functionalist theory of truth, 620, 723–749  
 as distinct from deflationism, 740–741  
 and pluralism about truth, 727–729, 733–735
- Gadamer, Hans-Georg, 163–164, 172, 175
- Geach, Peter, 660 (n. 1)
- Gibbard, Alan, 660 (n. 1)
- Gödel, K., 360 (n. 18), 626, 721 (n. 24)
- Goodman, N., 48–49, 172–173
- Grover, Dorothy, 425, 426, 486, 499 (nn. 3, 4), 505–526, 552 (n. 7), 557 (n. 49), 583, 584, 607 (n. 16), 677 (n. 6), 783 (n. 4). *See also* Prosentential theory of truth  
 claims prosentential theory of truth is neutral on substantive philosophical questions, 508, 509  
 defends prosentential theory of truth against objections, 512–526
- Gupta, Anil, 425, 427, 527–557, 566–568, 576 (n. 20), 725  
 claims deflationism cannot explain general facts involving truth, 531, 533–534, 538–539, 567–568  
 claims deflationism is committed to massive conceptual resources, 541–543  
 claims deflationists cannot explain the meaning of true, 544–552, 566

- Habermas, J., 281 (n. 21), 642  
 Haldane, J., 606 (n. 11)  
 Hale, Bob, 281 (n. 18), 606 (n. 11)  
 Hardy, J., 557 (n. 49)  
 Harman, G., 394, 414  
 Hazen, Allen, 557 (n. 49)  
 Heidegger, M., 172, 289–291, 295–316  
 Heil, J., 734  
 Hellman, G., 557 (n. 49)  
 Hempel, C., 52, 415 (n. 7)  
 Hill, Christopher, 554 (n. 19), 557 (n. 49)  
 Hills, David, 394  
 Hobbes, Thomas, 283 (n. 35), 692  
 Horgan, Terence, 13, 67–94, 620, 726, 746, 746 (n. 5)  
     and indirect correspondence theory of truth, 70–77  
 Hornsby, Jennifer, 617, 663–681, 702 (n. 1). *See also* Identity theory of truth  
     distinguishes identity theory from correspondence theory, 667–670  
 Horwich, Paul, 55, 426–428, 499 (n. 6), 512, 528, 537–539, 555 (n. 30), 557 (n. 49), 559–557, 584, 588, 598–599, 605, 606 (n. 6), 607 (n. 16), 608 (nn. 28, 30), 609 (n. 32), 616, 628, 632, 679 (n. 13), 680 (n. 17), 687, 709, 711, 718 (n. 5), 719 (nn. 6, 7), 721 (n. 23), 782 (nn. 2, 3), 783 (nn. 8, 11), 784 (n. 12)  
     Davidson's criticisms of, 632–636 (*see also* Davidson, Donald)  
     presents minimalist theory of truth, 559–561  
     replies to Davidson, 561–565 (*see also* Davidson, Donald)  
     replies to Dummett, 571 (*see also* Dummett, Michael)  
     replies to Field, 565–566 (*see also* Field, Hartry)  
     replies to Gupta, 556–557 (*see also* Gupta, Anil)  
     replies to Mark Richard, 571  
     replies to Soames, 568 (*see also* Soames, Scott)  
     Sosa's criticisms of, 650–653, 655–659  
     Hume, David, 624  
 Idealism, 99, 290, 670. *See also*  
     Absolute idealism; Metaphysical antirealism  
 Identity theory of truth, 617, 663–681, 683–704  
     and correspondence theory of truth, 667–670  
     David raises problems for, 689–690, 692, 694–695, 696–698  
     differs from deflationary theory of truth, 687–688  
     Hornsby's version of, 664–681  
     and realism about truth, 668, 675  
 Intellectual virtue in epistemology, 641, 649, 660  
 Internal (pragmatic) realism. *See* Putnam, Hilary  
 Intuitionism, 247–248  
 Irrationalism, 670. *See also* Antirealism; Idealism; Metaphysical antirealism  
 Jackson, Frank, 678 (n. 9), 744, 746, 747 (n. 13), 748 (n. 14)  
 James, William, 185, 186–187, 211, 211–228, 260, 444–445 (n. 12), 723  
     and agreement, 219–220  
     on reality, 215–218  
 Joachim, H. H., 99, 228 (n. 3)  
 Judgements as truth bearers, 99  
 Juridical truth, 735–738  
 Justification, 59–62, 641, 660  
 Kant, Immanuel, 313, 442, 604  
 Kaplan, D., 693  
 Kapus, Jerry, 523, 524 (n. 9), 525 (n. 23), 557 (n. 49)  
 Kirkham, Richard, 429 (n. 3), 524 (n. 8), 606 (n. 3), 607 (n. 15), 608 (nn. 25, 30)



- Klein, Peter, 746  
 Kokoszynska, M., 331  
 Kovach, Adam, 557 (n. 49)  
 Krausz, Michael, 281 (n. 22)  
 Kripke, Saul, 272, 274, 384, 396  
   (n. 17), 414, 418 (nn. 24, 25), 499  
   (n. 5), 501 (n. 16), 609 (n. 36), 692–  
   694, 716, 721 (n. 24)  
 Kuhn, Thomas, 271, 283 (n. 44), 286  
   (n. 61), 511
- Lance, Mark, 605, 606 (n. 8), 609  
   (n. 33)  
 Leeds, Stephen, 424, 485, 486, 499  
   (n. 3), 530, 531, 534, 535, 555  
   (n. 31), 599, 628, 630  
 Lehrer, Keith, 101  
 Lepore, Ernest, 280 (n. 9)  
 Leśniewski, S., 358 (n. 7)  
 Lewis, David, 71, 77, 278, 410, 414,  
   418 (n. 24), 497, 632, 729, 732,  
   743, 746 (n. 8)  
 Liar paradox, 324, 328 (n. 2), 339–  
   340, 414, 476–478, 479–480, 576  
   (n. 22)  
 Lycan, William, 605  
 Lynch, Michael, 65 (n. 3), 572, 607  
   (n. 14), 723–749, 782  
 Lyotard, F., 291
- MacIver, M., 557 (n. 49)  
 Mackie, J. L., 145  
 McDowell, John, 271, 272, 273–274,  
   617, 663–664, 670, 676 (n. 2), 679  
   (n. 15), 702 (n. 1), 702 (n. 3)  
 McGinn, Colin, 703 (n. 9)  
 McGrath, Matthew, 660 (n. 5)  
 Metaphysical antirealism, 48–49, 67–  
   70, 79–83, 399–400  
 Metaphysical realism, 11–12, 48–49,  
   55, 67–68, 77–79, 251–252, 618,  
   644, 679 (n. 15), 708, 709–710,  
   711, 712. *See also* Putnam, Hilary;  
   Realism  
 Millikan, Ruth, 279–280 (n. 4), 746  
   (n. 8)
- Minimalism, 260, 264  
 Minimalism (Horwich), 559–557,  
   633, 650–659, 782 (n. 2). *See also*  
   Horwich, Paul  
 Minimalism (Wright), 751–787. *See*  
   *also* Wright, Crispin  
   distinguished from deflationism, 751–  
   754  
   and pluralism about truth, 752–753,  
   759–761  
 Misak, C., 510, 523 (n. 1)  
 Moore, G. E., 444 (n. 6), 615, 625,  
   653–655, 661 (n. 8)  
   advocated primitivism about truth,  
   653–655  
   and identity theory of truth, 684–  
   685, 688–689, 702 (n. 1), 703  
   (n. 8)
- Nagel, E., 331  
 Nair, Ranjit, 557 (n. 49)  
 Neale, S., 677 (n. 6)  
 Negation equivalence (Wright), 756,  
   765, 766, 783 (n. 10)  
 Ness, A., 360 (n. 22), 361 (n. 29)  
 Neurath, O., 170, 415 (n. 7), 486  
 Nietzsche, F., 159, 319  
 Nonfactualism, 585–592
- O’Leary-Hawthorne, John, 513, 514–  
   515, 517–519, 521, 524 (nn. 12,  
   16)  
 Oppy, Graham, 513, 514–515, 517–  
   519, 521, 524 (nn. 12, 16), 678  
   (n. 9)
- Papineau, David, 279–280 (n. 4)  
 Peirce, C. S., 185–186, 193–210, 259,  
   761, 767–768, 784 (n. 16), 785  
   (n. 23)  
   defines truth, 206  
   and reality, 204–209  
 Performative theory of truth, 36–37,  
   423–434, 459–466, 572 (n. 1). *See*  
   *also* Strawson, P. F.  
 Pettit, Philip, 746, 746 (n. 7)

- Physicalism, 324, 375–377, 391, 395  
 (n. 13), 401–402, 411, 412, 415  
 (n. 9), 534–535, 645
- Pietroski, Paul, 605
- Pinker, S., 516
- Plantinga, Alvin, 65 (n. 8), 767–768,  
 785 (n. 26), 786 (n. 28)
- Plato, 623, 624
- Pluralism about truth, 618–621, 710–  
 714, 717–718  
 functionalist theory as version of,  
 727–729, 733–735  
 and Wright, 618–620, 725–726,  
 752–753, 759–761
- Popper, K., 415 (n. 7), 579, 606
- Postmodernism, 317–319
- Potrc, M., 91 (n. 3), 746
- Power and truth, 317–319
- Pragmatist theories of truth, 185–187,  
 188–190, 193–210, 211–228, 251–  
 258, 259–286, 641–642, 649, 761–  
 770. *See also* Dewey, J.; James,  
 William; Peirce, C. S.
- Primitivism about truth, 615–616, 661  
 (n. 8)  
 and Davidson, 625–628, 636, 661  
 (n. 8)  
 distinguished from deflationism, 655–  
 659  
 and Sosa, 653–651
- Propositions as truth bearers, 44–45,  
 424, 434, 473–474, 483, 563–564,  
 575 (n. 14), 634, 653, 678 (n. 10),  
 683, 685–687, 691–694, 701, 703  
 (n. 11), 715, 716, 732
- Prosentence, 425, 438, 506–507
- Prosentential theory of truth, 425–  
 426, 499–500 (nn. 4, 8), 505–526,  
 572 (n. 1), 583, 677 (n. 5), 783  
 (n. 4). *See also* Grover, Dorothy;  
 Brandom, Robert
- Pure disquotational truth (Field), 425,  
 484–488
- Putnam, Hilary, 57, 67, 71, 147, 167,  
 170–171, 173–174, 188–190, 251–  
 258, 260, 261, 280 (n. 4), 286  
 (n. 63), 417 (n. 23), 501 (n. 14),  
 535–536, 554 (n. 19), 555 (n. 24),  
 628, 635, 693, 705–722, 746 (n. 8),  
 764–770, 771, 783 (n. 6), 784 (nn.  
 17, 19), 785 (n. 23)
- and pluralism about truth, 618, 710–  
 714, 717–718, 726
- criticizes metaphysical realism, 251–  
 254, 709–710, 711
- defines metaphysical realism, 251
- and internal (pragmatic) realism, 49,  
 57, 81–82, 167, 189, 251–252,  
 256–258, 679 (nn. 14, 15)
- objections to views of, 57–64, 189,  
 764–770
- and truth as idealized rational  
 acceptability, 257, 642, 679 (n. 14),  
 761–764, 768–769, 771
- Quantificational theory of truth, 572–  
 573 (n. 1)
- Quasi-realism, 585
- Quietism, 272–275, 284 (n. 46)
- Quine, W. V., 74–75, 76, 93 (n. 18),  
 127, 327, 377, 390, 391, 398, 405,  
 407, 421, 429 (n. 3), 553 (nn. 8,  
 13), 554 (n. 15), 556 (n. 40), 628,  
 640, 668, 670–671, 676 (n. 3), 677  
 (n. 6), 708, 720 (n. 16)
- and disquotationalism, 424, 473–  
 481, 485, 486, 487, 491–492, 499  
 (n. 3), 512, 513–514, 529, 530,  
 531, 534, 630–632 (*see also*  
 Deflationism; Disquotationalism)
- on Tarski, 478–480
- Railton, Peter, 607 (n. 14)
- Ramberg, Bjorn, 283 (n. 36)
- Ramsey, F. P., 421, 429 (nn. 1, 6),  
 433–446, 484, 498, 499 (n. 7), 502  
 (n. 20), 638, 710, 732
- is not a deflationist, Field argues,  
 489–490
- Realism, 5, 9–14, 41–42, 45–47, 244–  
 246, 365, 399–400, 590–591, 643,  
 659, 675, 679 (n. 15), 708. *See also*

- Realism (cont.)  
 Correspondence theory of truth;  
 Metaphysical realism  
 defined in relation to truth, 9  
 objections to, 12–13, 51–57, 246–  
 249, 705–707, 708
- Realist conception of truth. *See* Alston,  
 W. P.; Correspondence theory of  
 truth; Realism
- Reduction, 378–384, 391–392
- Redundancy theory of truth, 232–237,  
 352–353, 421–422, 461–462, 572  
 (n. 1), 582, 649. *See also*  
 Deflationism; Ramsey, F. P.  
 Dummett's criticisms of, 232–237
- Reference, 68–70, 336, 367, 379–388,  
 389, 390, 391, 443, 468. *See also*  
 Denotation; Tarski, A.  
 causal theories of, 68, 253–255, 327,  
 384–385, 390–391, 404, 497, 603–  
 604  
 and primitive denotation, 367–368,  
 370, 389, 402–403, 404, 406  
 propositional (Ramsey), 434–435,  
 437–438, 443
- Relativity of truth, 131–132, 167–  
 168, 172–175, 186, 256–257
- Rescher, N., 124
- Rey, Georges, 605
- Rhees, R., 721 (n. 25)
- Richard, Mark, 569, 608 (n. 27)
- Robust vs. deflationary views of truth,  
 5. *See also* Correspondence theory  
 of truth; Deflationism
- Rorty, Richard, 52, 56–57, 189–190,  
 259–286, 291, 528, 680 (n. 17)  
 criticizes Crispin Wright, 265–275  
 on Davidson, 260–264  
 and pragmatist theory of truth, 259–  
 260, 641–642
- Rosenberg, G., 557 (n. 49)
- Rosenkranz, S., 782
- Rouse, J., 166
- Rumfitt, I., 676 (n. 4)
- Russell, Bertrand, 9–10, 17–24, 123,  
 147, 155 (n. 1), 436, 443 (n. 1), 444  
 (n. 7), 478, 485, 524, 625 (n. 8),  
 615, 653, 661 (n. 8), 668, 677 (n. 6),  
 684, 698, 699, 702 (n. 1)
- Rynin, D., 331
- Sainsbury, Mark, 726
- Satisfaction (Tarski), 325, 345–347,  
 382–383, 479
- Sayre-McCord, G., 606 (n. 11), 607  
 (n. 14)
- Schema T, 46, 73, 266, 323, 334–336,  
 356, 390, 393, 395–396 (n. 14),  
 413, 421, 475, 530, 541, 627–628,  
 660 (n. 3), 706, 728, 751, 782 (n. 1).  
*See also* Convention T; Equivalence  
 schema; Tarski, A.
- Schiller, F., 211, 226–227
- Schlick, M., 53, 123
- Searle, John, 285 (n. 59)
- Seligman, Jerry, 557 (n. 49)
- Sellars, W., 77, 280 (n. 4), 604
- Semantic ascent, 398–399  
 Quine on, 476
- Semantic conception of truth. *See*  
 Tarski, A.
- Sentence tokens as truth bearers, 43–  
 44, 369–370
- Sentence types as truth bearers, 42–43,  
 333–334, 424, 701, 703 (n. 12)
- Shoemaker, Sydney, 748 (n. 15)
- Shope, Robert, 786 (n. 29)
- Sinnott-Armstrong, Walter, 414
- Smith, Barry, 283 (n. 36)
- Smith, Michael, 678 (n. 9), 742, 784  
 (n. 15)
- Soames, Scott, 327–328, 387–418,  
 486, 497–498, 529, 557 (n. 49),  
 568, 575 (n. 18), 597, 607 (n. 14),  
 632, 635  
 on Tarski, 387–418, 497–498, 628
- Socrates, 623
- Sokal, Alan, 160
- Sosa, Ernest, 615, 616, 641–662  
 defends primitivism about truth,  
 653–660  
 on Horwich, 650–653, 655–659

- Statements as truth bearers, 26–38, 44, 434, 447–450
- States of affairs, 28–30, 454–456, 644, 669, 693–694, 703 (n. 7). *See also* Facts
- Sterelny, K., 607 (n. 13)
- Strawson, P. F., 10, 36–37, 273, 421, 423–424, 447–471, 499 (n. 3), 511. *See also* Performative theory of truth
- objects to facts as truth makers, 450–458
- on statements, 447–450
- Substitutionalism (Sosa), 651
- Substitutional quantification, 484, 486, 497, 498, 633, 644, 648, 651, 677 (n. 5)
- Superassertibility, 619, 737–738, 770–781, 786 (n. 33), 787 (nn. 34, 35, 36, 40). *See also* Wright, Crispin
- Supervenience, 659–660
- Szubka, T., 746
- Tarski, A., 261, 331–418, 421, 478–480, 512, 528, 573 (n. 1), 595–597, 615, 637, 678, 705, 706, 714, 718 (n. 4), 720 (nn. 16, 17), 721 (n. 24) and correspondence theory of truth, 327–328, 333–334, 348–349, 384, 628
- Davidson on, 625–630, 639
- and deflationism, 328, 400–401, 414, 478–480, 629–630
- Field on, 326–327, 365–396, 495–499, 501 (n. 16)
- and language, 326, 369–371, 400, 410–413, 498
- and material-adequacy condition, 323, 327, 334, 383 (*see also* Schema T)
- objections to, 369–388, 408–409, 417 (n. 23)
- and object language/metalanguage distinction, 341–343
- and physicalism, 324, 375–377, 391, 395 (n. 13), 401–402, 411, 412, 415 (n. 9)
- presents semantic conception of truth, 336–348
- Quine on, 478–480
- and satisfaction, 345–347, 382–383, 479
- and Schema T, 323, 334–335, 413, 356, 390, 393, 395–396 (n. 14)
- Soames on, 387–418
- Taylor, C., 291
- Tennant, Neil, 783 (n. 10)
- Thinkables as truth bearers (McDowell, Hornsby), 663–664, 670–674. *See also* Identity theory of truth
- Timmons, M., 88
- Truth. *See also* Correspondence theory of truth; Deflationism; Disquotationalism; Functionalism theory of truth; Identity theory of truth; Minimalism; Performative theory of truth; Pluralism about truth; Pragmatist theories of truth; Primitivism about truth; Prosentential theory of truth; Redundancy theory of truth; Tarski, A.
- distinction between concept of, and property of, 3, 50, 733, 753–754
- explanatory role of, 388–389, 398–399, 424, 518–523, 535–538, 656
- and justification, 59–62, 641, 649, 660, 760, 775–778
- and knowledge, 2, 641, 656–657, 674–675
- and meaning, 2, 408–409, 492–493, 539–542, 561–563, 602–604, 635–636, 670–671, 673–674, 679 (n. 13)
- need for concept of, 475–476, 485, 513–518, 582–583
- need for property of, 513–518, 535–538, 647, 756
- as normative, 230–231, 571–572, 577 (n. 25), 756–757
- Truth bearers, 9, 42–45, 332–333, 433–434, 473–475, 483, 676 (n. 3), 701

- Truth bearers (cont.)  
 beliefs as, 18, 21–23, 26, 125–126, 434, 652  
 eternal sentences as (Quine), 474  
 judgements as, 99  
 propositions as, 44–45, 424, 434, 473–474, 483, 563–564, 575 (n. 14), 634, 653, 678 (n. 10), 683, 685–687, 691–694, 701, 703 (n. 11), 715, 716, 732  
 sentence tokens as, 43–44, 369–370  
 sentence types as, 42–43, 333–334, 424, 701, 703 (n. 12)  
 statements as, 26–38, 44, 434, 447–450  
 thinkables as, 663–664, 670–674  
 utterances as, 483–484, 500 (n. 9), 652
- Truth conditions, 483–484, 602–604, 635–636
- Truth makers, 9, 698–700. *See also* Facts; States of affairs
- T-schema (Alston), 46. *See also* Schema T
- Untruth (Heidegger), 307–312. *See also* Falsehood
- Utterances as truth bearers, 483–484, 500 (n. 9), 652
- Vagueness, 83–87, 91 (n. 9)
- Verification, 213–214
- Verificationism, 187–188, 247–248, 324, 492, 579, 618, 705–707, 710–714
- Von Juhos, B., 361 (nn. 25, 26)
- Walker, Ralph C. S., 101, 123–158, 747 (n. 11), 747 (n. 13). *See also* Coherence theory of truth  
 distinguishes pure from impure coherence theories of truth, 146–147, 151–155  
 objects to pure coherence theory of truth, 147–149
- Wallace, John, 394
- Warranted assertibility, 71, 755, 756, 757, 770–771
- Westmoreland, Robert, 746
- Wiggins, David, 258, 269, 270
- Williams, Bernard, 276
- Williams, Michael, 52, 177, 279 (n. 2), 281 (n. 21), 429 (n. 5), 528, 529, 536–537, 539, 553 (n. 15), 555 (n. 25), 629, 630, 646–647, 746
- Wilson, Kent, 524 (n. 5)
- Wilson, Mark, 556 (n. 46), 557 (n. 49)
- Wittgenstein, L., 160, 173, 236, 249, 273, 274, 277, 284 (n. 52), 485, 604, 609 (n. 36), 635, 698, 699, 710, 720 (nn. 14, 15, 18, 19, 20), 721 (nn. 22, 24, 25, 26), 722 (n. 27), 734
- Putnam's account of views on truth of, 714–718
- Wolenski, Jan, 718 (n. 4)
- Wollheim, R., 157 (n. 18)
- Wright, Cory, 746
- Wright, Crispin, 71, 77, 92 (n. 11), 157 (n. 17), 259, 606 (n. 11), 607 (n. 14), 669, 737, 746, 746 (nn. 3, 5, 6, 7), 751–787. *See also* Minimalism (Wright); Superassertibility  
 and analytic theory of concept of truth, 730–732, 759–761  
 objects to deflationism, 754–759  
 and pluralism about truth, 618–620, 725–726, 752–753, 759–761  
 and pragmatism, 762–770  
 Rorty's criticisms of, 265–279  
 and superassertibility, 619, 737–738, 770–781, 787 (nn. 34, 35, 36, 40)
- X, Malcom, 160